

1 **Assessment of Planetary Boundary Layer parametrizations and urban heat**
2 **island comparison: Impacts and implications for tracer transport**

3 Israel Lopez-Coto*

4 *National Institute of Standards and Technology, Gaithersburg, MD*

5 Micheal Hicks[†]

6 *National Weather Service, National Oceanic and Atmospheric Administration, Sterling, VA*

7 Anna Karion

8 *National Institute of Standards and Technology, Gaithersburg, MD*

9 Ricardo K. Sakai

10 *Howard University, Washington, DC*

11 Belay Demoz

12 *Department of Physics, University of Maryland, Baltimore County, Baltimore, MD*

13 Kuldeep Prasad and James Whetstone

14 *National Institute of Standards and Technology, Gaithersburg, MD*

15 *Corresponding author address: National Institute of Standards and Technology, Gaithersburg, MD

16 E-mail: israel.lopezcoto@nist.gov

¹⁷ †Current Affiliation: National Institute of Standards and Technology, Gaithersburg, MD

ABSTRACT

18 Accurate simulation of planetary boundary layer height (PBLH) is key to
19 greenhouse gas emission estimation, air quality prediction and weather fore-
20 casting. This manuscript describes an extensive performance assessment
21 of several Weather Research and Forecasting (WRF) model configurations
22 where novel observations from ceilometers, surface stations and a flux tower
23 were used to study their ability to reproduce planetary boundary layer heights
24 (PBLH) and the impact that the urban heat island (UHI) has on the mod-
25 eled PBLHs in the greater Washington, D.C. area. In addition, CO₂ mea-
26 surements at two urban towers were compared to tracer transport simulations.
27 The ensemble of models used 4 PBL parameterizations, 2 sources of initial
28 and boundary conditions and 1 configuration including the building energy
29 parameterization (BEP) urban canopy model. Results have shown low biases
30 over the whole domain and period for wind speed, wind direction and temper-
31 ature with no drastic differences between meteorological drivers. We find that
32 PBLH errors are mostly positively correlated with sensible heat flux errors,
33 and that modeled positive UHI intensities are associated with deeper mod-
34 eled PBLs over the urban areas. In addition, we find that modeled PBLHs
35 are typically biased low during nighttime for most of the configurations with
36 the exception of those using the MYNN parametrization and that these biases
37 directly translate to tracer biases. Overall, the configurations using MYNN
38 scheme performed the best, reproducing the PBLH and CO₂ molar fractions
39 reasonably well during all hours, thus opening the door to future nighttime
40 inverse modeling.

41 **1. Introduction**

42 Turbulent mixing drives the transport of mass, heat and momentum in the planetary boundary
43 layer (PBL) (Stull (1988)) and, therefore, numerical weather prediction (NWP) models need to
44 include PBL parametrizations to ensure that this phenomenon is properly represented. In addition,
45 atmospheric transport models rely strongly on the proper representation of the PBL by the NWP
46 model driving them to properly account for the mixing of pollutants. These transport models are
47 fundamental tools for air quality prediction as well as for the inference of trace gas (pollutant or
48 greenhouse gas) sources using top-down approaches (Nisbet and Weiss (2010)).

49 Many different PBL schemes are available; they differ from each other by the vertical mixing
50 formulation (local vs. non-local) and the closure order. Local schemes only consider adjacent ver-
51 tical levels in the fluxes computations, while nonlocal schemes take into account multiple levels,
52 often from the surface up to the estimated PBL height, in representing the fluxes through the PBL.
53 In addition, PBL schemes are coupled to the surface layer parametrizations, that generally are not
54 interchangeable, and strongly influence the near surface variables and PBL mean properties, (Shin
55 and Hong (2011)).

56 Recent studies have looked at different PBL schemes with the focus of atmospheric transport
57 modeling in mind. For example, Angevine et al. (2012) and Feng et al. (2016) studied the per-
58 formance of different PBL schemes in the Weather Research and Forecasting (WRF) model along
59 with other physics options for the CalNex-2010 campaign (late spring, 2010). Kretschmer et al.
60 (2012, 2014) compared the impact of two PBL schemes on CO₂ transport over Europe and eval-
61 uated them with radiosondes during late summer. Sarmiento et al. (2017) studied the behavior
62 of PBL schemes and their interactions with Land Surface models and the land use representation
63 over Indianapolis (Indiana, USA) for a month in late winter and a month in summer. Lian et al.

64 (2018) studied WRF PBL schemes and their impacts on CO₂ transport for a month in winter over
65 Paris (France) area. Díaz-Isaac et al. (2018) did a comparison of multiple WRF physics schemes
66 for a summer month in the Midwest of the United states. Over the Washington DC - Baltimore
67 (Maryland, USA) area, WRF PBL schemes were also evaluated as part of the DISCOVER-AQ
68 campaign (Hegarty et al. (2018)) during July 2011. These studies demonstrate that there is much
69 interest in finding the best performing configuration for WRF so that the errors introduced in trace
70 gas transport are minimized. However, the results obtained are somewhat dependent on the region
71 and period studied, the observations used for verification, the methods applied to derive PBLH and
72 the WRF version.

73 PBLH observations are not very common. Their availability is sparse, in space and time, and
74 rely strongly on operational radiosondes that sample the PBL only twice a day. This lack of
75 measurement data limits understanding of PBL dynamics and validation studies, and therefore pa-
76 rameterization development. The introduction of new measurement techniques for mixing height,
77 such as those based on ceilometers and particle Lidars, has the potential to be a game changer for
78 model validation due to the greater temporal coverage and resolution that they provide. In the last
79 few years, Lidar observations and ceilometers have been used to evaluate WRF simulations (Ware
80 et al. (2016); Feng et al. (2016); Hegarty et al. (2018)).

81 Impervious urban surfaces are characterized by lower albedo, lower specific heat capacity, higher
82 thermal conductivity and much smaller rainfall retention than rural surfaces (Oke (1982)). These
83 properties cause higher Bowen ratios (larger sensible heat fluxes and lower latent heat fluxes) and
84 surface temperatures in the urban landscape and, therefore, induce perturbations in the wind, air
85 temperature, water vapor content as well as in the boundary layer height, (Angevine et al. (2003);
86 Zhang et al. (2009, 2011)). Understanding how the meteorological models reproduce this feature
87 is also essential for atmospheric transport.

88 The WRF model undergoes continuous development with two releases per year as new mea-
89 surements and techniques become available, therefore new comparisons and testing are needed.
90 In addition, it is clear from previous studies that there is no single configuration that works best
91 under all circumstances and validation for specific areas and periods are required.

92 In this work, we intend to better understand the performance of eight configurations of WRF
93 over the Washington DC/Baltimore area during winter, to uncover similarities and differences in
94 PBL parametrizations regarding PBLH and urban heat island related variables and the impacts on
95 tracer transport with the aim of identifying the best performing configuration for the purpose of
96 greenhouse gas (GHG) inverse modeling in the North East Corridor - Baltimore / Washington DC
97 test bed (Lopez-Coto et al. (2017a)). In Section 2, the eight model configurations as well as the
98 surface stations, CO₂ measurements, flux tower and ceilometers used for comparison are described.
99 In Section 3, the model performance is presented as well as an analysis between the heat island
100 produced by each configurations and how it impacts on PBLH. In Section 4, implications of our
101 findings on tracer transport and inverse modeling are discussed. Last, in Section 5, the main
102 conclusions obtained are highlighted.

103 **2. Methods**

104 *a. Observational data*

105 1) SURFACE STATIONS

106 The Integrated Surface Database (ISD) consists of global hourly and synoptic observations from
107 more than 100 original data sources that collectively archived hundreds of meteorological vari-
108 ables. It is compiled by the NOAAs National Climatic Data Center (NCDC) and accessible
109 through the web (<https://www.ncdc.noaa.gov/isd>). The primary data sources include the Auto-

110 mated Surface Observing System (ASOS), Automated Weather Observing System (AWOS), Syn-
111 optic, Airways, METAR, Coastal Marine (CMAN), Buoy, and various others, from both military
112 and civilian stations including both automated and manual observations (Smith et al. (2011)).
113 More than 14,000 active stations worldwide are updated daily in the database. As described in
114 Smith et al. (2011), ISD contains 54 quality control (QC) algorithms, which serve to process
115 each data observation through a series of validity checks, extreme value checks, internal (within
116 observation) consistency checks, and external (versus another observation for the same station)
117 continuity checks. For the month of February 2016 and the domain of interest, six ISD surface
118 stations had data with the highest level of quality control flag, Figure 1.

119 2) CO₂ MEASUREMENTS

120 Three towers equipped with Cavity Ring Down Spectrometers are used to measure CO₂. The
121 sites, NDC, HAL and BUC are located near Washington, D.C., Baltimore, M.D., and a more back-
122 ground area in the Delmarva peninsula about 100 km away from the urban centers respectively.
123 Further details about the stations, calibration and quality control can be found in Karion et al.
124 (2020).

125 CO₂ enhancements were computed subtracting from each hourly observation, the measurements
126 at the background tower (BUC) similarly to other work in urban areas (Lauvaux et al. (2016).)

127 3) CEILOMETERS AND PLANETARY BOUNDARY LAYER HEIGHT (PBLH) RETRIEVAL

128 Two Vaisala ceilometers were used to derive PBLHs during the period of interest: the CL-31 at
129 the National Weather Service (NWS) Sterling Field Support Center (SFSC) in Sterling, VA and
130 the CL-51 at Beltsville, MD (HUBV) (Figure 1)

131 Both ceilometers use an InGaAs laser diode with a 910 nm wavelength. They use a single lens
132 optics system where the inner part of the lens is employed for transmitting and the outer part for
133 receiving light. This system provides a good overlap of the transmitter and the receiver field-of-
134 view over the whole measuring range, conferring an improved near-range performance compared
135 to two lens systems and allows reliable detection of very low nocturnal stable layers below 200 m.
136 The latest model, the CL-51, is equipped with a larger lens and a more powerful laser transmitter
137 module. These improvements increase the reporting range and signal-to-noise ratio.

138 Thirty-minute averaged two-way attenuated backscatter profiles with a vertical resolution of 20
139 m from the surface are processed to derive PBL heights using the Wavelet Covariance Transform
140 (WCT; Davis et al. (2000); Compton et al. (2013)) method for unstable/neutral conditions and
141 the Hybrid-Lowest for stable conditions (Hicks et al. (2015)). The stability was determined in
142 base to the averaged low-altitude Bulk Richardson values (below 0.2 km); values lower than -
143 0.01 were considered unstable, near neutral if between or equal to -0.01 and 0.01, and stable if
144 greater than 0.01. The Hybrid-Lowest method is a combination of the WCT method and the Error
145 Function-ideal profile (ERF; Steyn et al. (1999)) method. They are combined such that the WCT
146 method detects the significant gradient layers and the ERF method determines which of the layers
147 correspond to PBL height below significant elevated aerosol layers. We note that the Hybrid-
148 Lowest algorithm expects there to be a residual layer at night and attempts to locate the PBL
149 height beneath it. In addition, a height constraint based on the lifting condensation level (LCL) is
150 applied for both techniques. Details of the methods can be found in Hicks et al. (2015, 2019).

151 The PBLHs derived from the ceilometers, as described above, were manually filtered after visual
152 inspection by removing those corresponding to rain events and other apparent artifacts on the
153 backscatter signals such as a dirty lens or malfunctioning hardware leading to an acceptance rate
154 of 77.4 % (N=1100) for SFSC and 73.4 % (N=1043) for HUBV.

155 In Hicks et al. (2015, 2019) the consistency rate (Co) parameter was defined as the percentage
156 of the ceilometer PBLH observations that measured within ± 300 m when compared to radioson-
157 des. The consistency rate reported in these previous works for the PBLH retrieval methods em-
158 ployed was 65 % for unstable conditions and 74 % for stable conditions. These comparisons were
159 performed in the context of the NWS CL31 Planetary Boundary Layer project, (Atkinson et al.
160 (2017)). Here we use the consistency rate (Co) parameter to compare the simulated PBLH to the
161 ceilometer observations.

162 4) FLUX TOWER

163 At HUBV, a micrometeorological tower has been collecting micrometeorological parameters
164 since 2006. The campus is located in a complex suburban/rural/industrial landscape, however, the
165 campus itself has minimal urban development, and it is principally covered by a mix of deciduous
166 (maple and mixed oak) and coniferous (mainly Virginia Pine) trees. At the tower, fast response
167 instruments measure variables such as wind speed, temperature (CSAT, Campbell Scientific), wa-
168 ter vapor, and CO₂ concentrations (LI7500, LICOR Inc) at 31.5 m above ground level (~ 15 m
169 above the canopy). Before the eddy covariance technique (Stull (1988)) is used to estimate turbu-
170 lent fluxes, a flow rotation to the wind field is applied (McMillen (1988)). Also, scalar fluxes are
171 corrected due to density fluctuations (Webb et al. (1980)).

172 *b. Model configurations*

173 1) METEOROLOGICAL MODEL

174 Simulations for the month of February 2016 were conducted with the Weather Research and
175 Forecasting (WRF) model. February was selected because is a representative month of winter
176 in the study area. In addition, inverse modeling studies are carried out mostly during winters

177 so that biogenic activity is small and cause little influence on CO₂ estimated fluxes. The Ad-
178 vanced Research WRF (ARW) core uses fully compressible, non-hydrostatic Eulerian equations
179 on an Arakawa C-staggered grid with conservation of mass, momentum, entropy, and scalars (Ska-
180 marock et al. (2008)).

181 Two datasets were tested as initial and boundary conditions: North America Regional Reanalysis
182 (NARR) three hourly data (Mesinger et al. (2006)) and High Resolution Rapid Refresh (HRRR)
183 model hourly analyses (Benjamin et al. (2016)) following Blaylock et al. (2017), both provided
184 by the National Center for Environmental Prediction (NCEP). As in Lopez-Coto et al. (2017b),
185 the model was configured with 3 nested domains (with feedback) of 9, 3 and 1 km horizontal
186 resolution respectively. However, for the case of HRRR, only two domains were used being 3 and
187 1 km horizontal resolution. 60 vertical levels with monotonically increasing thickness from the
188 surface resulted in 34 levels below 3 km for better boundary layer representation. Adaptive time
189 step was selected with a Courant-Friedrichs-Lewy (CFL) criterion of 1. The RRTMG radiation
190 scheme, (Mlawer et al. (1997)), Thompson microphysics scheme, (Thompson et al. (2004, 2008)),
191 Noah land surface model, (Chen and Dudhia (2001)) and the Kain-Fritsch cumulus scheme, for
192 the 9 km domain only, (Kain (2004)) were used and kept constant across configurations.

193 Four PBL schemes were compared, three local schemes and one non-local scheme: 1) YSU is
194 a nonlocal, first order closure scheme. It includes a countergradient correction term in the down-
195 gradient diffusion and represents the entrainment explicitly (Hong et al. (2006)). Later on, Hong
196 (2010) removed the counter-gradient flux terms and included other changes for stable boundary
197 layers. The PBL height in the YSU scheme is determined from the Rib method calculated from the
198 surface to the top of the PBL. A threshold value of zero is used for stable cases, while 0.25 is used
199 for unstable conditions. 2) QNSE is a local, 1.5- order local closure, scheme (Sukoriansky et al.
200 (2005)). It is intended to account for wave phenomena within stable boundary layers. The QNSE

201 theory is valid for stable stratification and weakly unstable conditions. The PBLH is diagnosed
202 based on a TKE threshold. 3) BouLac is also a local, 1.5-order local closure scheme including
203 a prognostic equation for TKE, (Bougeault and Lacarrere (1989)). It is designed for use with
204 the BEP (Building Environment Parametrization) multi-layer, urban canopy model (Martilli et al.
205 (2002); Salamanca et al. (2011a,b)). BouLac diagnoses PBL height as the height where the vir-
206 tual potential temperature exceeds the surface virtual potential temperature by 0.5 K. Here we use
207 this scheme with and without the BEP parametrization. 4) MYNN is a local scheme (Nakanishi
208 and Niino (2004, 2006)). In particular, we tested the 1.5- order closure scheme (MYNN2). The
209 expressions of stability and mixing length are based on the results of large eddy simulations rather
210 than on observations. In recent years, MYNN has undergone extensive development, including the
211 addition of BouLac mixing length in the free atmosphere, changing the turbulent mixing length to
212 be integrated from the surface to the top of the boundary layer plus a transition layer depth, the
213 addition of a scale-aware mixing length following Ito et al. (2015) and the addition of an eddy
214 mass-flux option (Angevine et al. (2018); Olson et al. (2019)) that confers to this scheme some
215 non-local characteristics as well. For the PBL height diagnosis, a hybrid method is used, which
216 blends a theta-v-based definition in neutral/convective boundary layer and a TKE-based definition
217 in stable conditions. We tested here MYNN with and without the eddy mass-flux option.

218 The land-use classification plays a role in the model since it determines the values for the surface
219 properties as the roughness length, albedo and heat capacity which are important for the surface
220 energy balance and heat and momentum fluxes to the atmosphere. Here we tested two datasets
221 available in WRF: The USGS dataset and the more up to date NLCD 2011. The main difference
222 between these datasets, as concerns this work, is the representation of the urban land use; in
223 the USGS, only one urban category is defined while in the NLCD, four categories exist going
224 from developed open space to developed high intensity (Figure 1). For the configuration using

225 the Building Energy Parameterization (BEP) multilayer Urban Canopy Model (UCM) a modified
226 version of USGS was used that included 3 urban categories taken from the NLCD dataset where
227 the developed open space and low intensity categories were added together. In addition, the BEP
228 specific parameters for Washington, DC, and Baltimore, MD, cities were taken from the NUDAPT
229 dataset already included in the WRF data distribution.

230 All the options described above total to eight different configurations that were tested here (Table
231 1).

232 Model-data comparison was performed similarly for each data source. Hourly (or half-hourly)
233 measurements for each station were compared to model predictions extracted at the location and
234 time of each observation. Then, daily cycles, bias (model - observations), standard deviation of the
235 differences and percentiles were computed for all stations together. In the case of the ceilometers,
236 metrics for each ceilometer are provided as well as both combined.

237 2) TRACER TRANSPORT MODEL

238 The CO₂ transport was simulated similarly to Lopez-Coto et al. (2017a). We used the Stochastic
239 Time-Inverted Lagrangian Transport model (STILT; Lin (2003); Nehrkorn et al. (2010)), driven
240 by meteorological fields generated by four of the configurations described above (MYNN, YSU,
241 QNSE, BOUL+UCM). Five-hundred particles were released from both urban sites (NDC and
242 HAL) hourly, and were tracked as they moved backwards in time for 24 h. The footprint was
243 calculated from the particle density and residence time in the layer that sees surface emissions,
244 defined as 0.5 PBLH (Gerbig et al. (2003)) and then convolved with CO₂ fluxes provided by
245 ACES inventory (Gately and Hutyra (2017)).

246 3. Results

247 a. Surface variables

248 Overall, over the month of February, the temperature bias ranged from -0.92 K for MYNNe+nlcd
249 to 1.96 K for BOUL+UCM while standard deviation ranged from 1.60 K for YSU to 2.04 K for
250 MYNNe and YSU+NARR. Wind speed bias was negative for all but BOUL, ranging from -1.12
251 m/s for MYNNe+nlcd to 0.69 m/s for BOUL. The standard deviation ranged from 1.70 m/s for
252 YSU and QNSE to 2.16 m/s for BOUL+UCM. Wind direction bias ranged from -5.03° for QNSE
253 to 6.43° for BOUL+UCM while standard deviations did from 36.62° for BOUL to 47.12° for
254 MYNN (Table 2). It is interesting that by looking at YSU and YSU+NARR, it seems that the
255 HRRR driver provided better results; however, by looking at MYNN and MYNNe, the conclusions
256 would be the opposite. It is worth noting that due to the circular nature of the wind direction,
257 differences larger (lower) than 180° (-180°) were measured in the opposite direction, for example,
258 if the model had a wind direction of 175 while the observations were at -160, the difference is
259 equal to -25, not 335. This step removes the fat tails of the distribution and makes them much
260 more Gaussian and hence gains significance for the mean and standard deviation calculated here.

261 In the overall statistics, the urban canopy model decreased the performance of the BouLac
262 parametrization for the 3 surface variables analyzed here making it too warm and more variable
263 regarding wind speeds and direction errors. In addition, the MYNN scheme showed the largest
264 wind direction error variability, especially when it was driven by HRRR. On the other hand, in-
265 cluding the eddy mass flux option in MYNN had a positive impact on wind speed and direction
266 but using the NLCD dataset caused the model to be colder and reduced the wind speeds.

267 The daily cycle of the temperature differences (Fig 2a) reflects that BOUL was too warm during
268 nights while during the day the median bias was close to zero. The addition of the UCM did

269 however increase the bias during the day as well. QNSE was colder during nighttime while for
270 MYNN and MYNNe the median temperature bias was similar during day and night. The inclusion
271 of NLCD caused the temperatures to decrease during day time. The wind speed errors had a
272 clear daily cycle for BouLac, with winds being too strong during night time (Fig 2b). The UCM
273 corrected this bias at the cost of increasing the (negative) bias during daytime. For the rest of
274 configurations, the bias was slightly more negative during daytime but not as marked as for BOUL.
275 No significant cycle was observed for the wind direction errors for any configuration (Fig 2c).

276 *b. Planetary Boundary Layer Heights*

277 1) OBSERVED PLANETARY BOUNDARY LAYER HEIGHTS

278 Figure 3 shows the daily cycle observed at the two ceilometer locations for the month of Febru-
279 ary 2016. The observed PBLHs are in good agreement with previous climatological results (Seidel
280 et al. (2012)), and more specifically with the results published for the area under study by Hegarty
281 et al. (2018).

282 Beltsville (HUBV) shows typically higher PBL depths during the day as well as sharper transi-
283 tions during the morning and especially during the evening as compared to Sterling (SFSC). This is
284 likely due to the more urban surroundings for that location and the typical westerly flows dominant
285 over the region that likely transport air masses with deeper PBL from the denser Washington DC
286 metro area adjacent to this location (Angevine et al. (2003); Zhang et al. (2009, 2011)). However,
287 we also note that the differences between ceilometers between both locations might have played a
288 role as well on the observed differences.

289 2) PERFORMANCE OF MODEL CONFIGURATIONS

290 Table 3 shows the overall statistics for the PBLH differences for the eight tested configurations
291 using both ceilometers together and for each ceilometer independently. For both ceilometers,
292 the consistency rate ranged from 52 % for QNSE and BOUL to 60 % for MYNN. The mean
293 values were higher than the median values indicating the differences were skewed to higher values.
294 In the extreme case, QNSE provided the largest values for all the statistics estimators but the
295 lowest consistency parameter. For each ceilometer, the results resemble the global values, showing
296 slightly better model performance for Beltsville (HUBV) than for Sterling (SFSC). Reasons for
297 that are not clear but it is possible that the Vaisala CL-51 ceilometer at HUBV, which has a better
298 signal-to-noise ratio than the CL31 at SFSC, might have achieved a higher quality PBLH retrieval.
299 Overall, the three variants of MYNN provided the best consistency rate and the lowest standard
300 deviation followed closely by YSU.

301 To better understand the performance of each model configuration, we analyzed both the daily
302 cycle of the differences (Fig. 2d) and the daily cycle of the relative differences, Figure 4. Noc-
303 turnal PBLH bias is typically smaller than daytime values. Although in relative terms, they are
304 comparable, or larger at night due to the typically low measured nocturnal PBLH values.

305 BOUL provided the lowest nocturnal PBLH values, reaching median bias between -50 % and
306 -60 % of the observed values, followed by YSU (- 50%). During daytime, both BOUL and YSU
307 performed much better providing median relative bias close to zero. The inclusion of the urban
308 canopy parameterization increased PBLH values, slightly improving the BouLac schemes noctur-
309 nal performance but at the cost of an increased daytime bias. The QNSE configuration gave the
310 largest PBLH values during the day, followed by BOUL+UCM. QNSE performed better during
311 nighttime but still over-predicted the PBLH. It also showed the largest IQR. On the other hand,

312 MYNN performed well during most hours, slightly under-predicting PBLH during the evening.
313 The usage of NARR driver data did however decrease the MYNN performance during the night
314 causing a slight underestimation of the PBLH during these hours. This configuration had also
315 the eddy mass flux option activated, however, the decreased nighttime performance cannot be at-
316 tributed to it because this option only gets triggered during convective situations. The inclusion
317 of the NLCD dataset had a positive impact on the prediction, causing MYNN+nlcd to be nearly
318 unbiased for all hours. Interestingly, most of the configurations showed a noticeable PBLH drop at
319 18 EST probably coinciding with the evening transition. It is not clear however whether this result
320 is caused by a too quick evening transition in the models or a problem on the retrieved PBLH in
321 this complex situation. Overall, MYNN produced the best predictions of the PBLH for all hours.

322 *c. Surface fluxes*

323 Table 4 shows the sensible and latent heat flux errors at Beltsville (HUBV). Sensible heat flux
324 bias ranged from -23 W m^{-2} for MYNN to 12 W m^{-2} for MYNN+nlcd while the standard devi-
325 ation ranged from 86 W m^{-2} for YSU+NARR to 97 W m^{-2} for BOUL+UCM. For the latent heat
326 flux, the bias ranged from -11 W m^{-2} for MYNN+nlcd to 21 W m^{-2} for QNSE. The standard
327 deviation ranged from 57 W m^{-2} for BOUL+UCM to 65 W m^{-2} .

328 The daily cycle of the sensible heat flux differences (Fig 2e) shows that all the configurations
329 are nearly unbiased from 17 EST to 7 EST with the exception of MYNN+nlcd, which shows a
330 slight positive bias during those hours. During daytime, the model performance is more variable:
331 YSU, YSU+NARR, MYNN, BOUL and QNSE are nearly unbiased during the morning while
332 showing negative bias during the afternoon; MYNN shows negative bias during all daytime hours,
333 being the largest in the afternoon; BOUL+UCM presents positive bias during all daytime, being
334 the largest during the late morning and MYNN+nlcd is nearly unbiased during these hours.

335 The daily cycle of the latent heat flux differences (Fig 2f) shows a similar behavior during non-
336 daylight hours (17 - 7 EST) as in the previous case with very little to no bias for all the schemes.
337 However, in this case MYNNe+nlcd and BOUL+UCM are the best performing configurations
338 with almost zero bias while the rest show a slightly positive bias. During daytime hours, all
339 the configurations show a positive bias that is the largest close to noon, with the exception of
340 MYNNe+nlcd which has a negative bias.

341 Figure 5 presents a scatter plot of the mean daily cycle of PBLH differences vs. the sensi-
342 ble heat flux differences. As expected, all configurations show a positive correlation between
343 the two, with the exception of BOUL+UCM. However, both the magnitude of the dependence,
344 as measured by the slope of a linear model, and the intensity of the correlation, as measured by
345 the Pearson correlation coefficient, differs between configurations. BOUL and YSU exhibit the
346 largest slope of them all, followed by MYNNe, QNSE, YSU+NARR, MYNN, MYNNe+nlcd
347 and BOUL+UCM, which is the only one with negative slope. The correlations are between
348 0.66 and 0.8 for MYNN, YSU+NARR, MYNNe, YSU, QNSE and BOUL, but below 0.35 for
349 MYNNe+nlcd and BOUL+UCM. This analysis shows that when the model underestimates the
350 sensible heat flux, the PBLH tends to be underestimated as well. This is true for all the configura-
351 tions but BOUL+UCM and to a lesser extent for MYNNe+nlcd.

352 *d. Urban Heat Island Effect*

353 The Urban Heat Island (UHI), computed here as the difference between the area-averaged sur-
354 face skin temperature (TSK) for the urban area and the non-urban area, is reproduced similarly
355 by all configurations being about 2 - 3 K during nights with a peak during early evening (Fig 6a).
356 During the daytime, the median values are close to zero for all configurations but BOUL+UCM
357 which showed UHI intensities over 2 K for these hours. As shown by Basara et al. (2008), and by

358 comparison to the rest of the models, the values shown by BOUL+UCM are rather large. Over-
359 all, all models reproduced the larger sensible heat fluxes in the urban areas as expected (Wood
360 et al. (2013)). However, QNSE and BOUL+UCM showed the largest contrast between urban and
361 non-urban sensible heat fluxes (ΔHFX) of all configurations being the largest for the latter (Fig
362 6c). In addition, the peak was at least three hours earlier than for the rest of configurations. The
363 usage of NARR driver data in MYNN caused the sensible heat flux differences between urban and
364 non-urban areas to increase while the addition of the NLCD land cover dataset had the opposite
365 effect.

366 On the other hand, BOUL+UCM had the smallest difference between urban and non-urban latent
367 heat fluxes (ΔLE) while QNSE had the largest (Fig 6d). The median Bowen ratio over urban areas
368 during daytime was between 5 to 15 times larger than those for the non urban areas for most of
369 the models, with MYNNe+nlcd being the smallest and QNSE the largest (Fig 6b). These values
370 are within the range of observed values but on the large side of the typical ones (Oke (1982)).
371 However, BOUL+UCM showed two peaks, at 8 a.m. and 4 p.m. (EST), with values of up to 20
372 times those of the non urban areas. This feature is not seen in any other configuration and can be
373 attributed to the UCM since the BOUL configuration without UCM behaved similarly to the rest
374 of configurations.

375 QNSE and BOUL+UCM had consistently the largest PBLH difference between the urban
376 and non-urban areas ($\Delta PBLH$) while MYNNe+nlcd had the lowest (Figure 7). QNSE and
377 BOUL+UCM also had positive $\Delta PBLH$ during daytime while the rest of configurations had a
378 median value close to zero during these hours. As with the UHI, the maximum differences were
379 simulated during the early evening, about 17 - 18 EST depending on the configuration. The us-
380 age of NARR driver data as well as the inclusion of the eddy mass flux option in MYNN caused
381 the median PBLH differences to decrease. For YSU, $\Delta PBLH$ was the most different of all con-

382 figurations having a median value near zero during nighttime but with the distribution skewed to
383 negative values indicating that in many occasions the PBL was deeper in the non-urban areas than
384 in the urban areas. Attending to the results published by Godowitch et al. (1985) and Angevine
385 et al. (2003) and by comparison with the rest of models, this result is not expected and seems odd.
386 The reasons for this are not clear because neither the UHI nor the ΔHFX showed a cycle that could
387 suggest this type of behavior.

388 To understand the relationship between the PBLH differences between the urban and non-urban
389 areas and the Urban Heat Island (UHI) intensity, Figure 8 shows a scatter-plot for all simulated
390 values for the month of February along with the slope of a linear model between the two variables
391 for each model configuration. Overall, negative UHI intensities resulted in deeper PBLs over the
392 non-urban areas while positive UHI intensities were associated with deeper PBLs over the urban
393 areas, as expected, (Godowitch et al. (1985); Angevine et al. (2003)). This relationship shows a
394 somewhat linear trend where larger UHI values resulted in larger PBLH differences between urban
395 and non-urban areas for all configurations but YSU. Slopes ranged from -1.8 ± 2.1 m/K for YSU
396 to 72.7 ± 2.4 m/K for QNSE. BOUL and BOUL+UCM had higher slopes than MYNN while the
397 inclusion of the NLCD dataset reduced the slope considerably from 47.2 to 29.7 m/K. Correlation
398 coefficients ranged from 0.02 and 0.06 for YSU and YSU+NARR respectively to 0.65 and 0.64
399 for BOUL and MYNN+nlcd respectively. The rest of the configurations also had correlation
400 coefficients larger than 0.5, except for BOUL + UCM, which had a correlation of 0.4.

401 The near zero correlation coefficient and slope showed by YSU and YSU+NARR is caused by
402 the large hysteresis shown in the median cycle of these two variables for both configurations (Fig-
403 ure 9). During night and until late morning, the median UHI intensity decreases while the PBLH
404 difference between urban and non-urban areas slightly increases. This behavior is the opposite to
405 the rest of configurations and previously published works (Spangler and Dirks (1974); Godowitch

406 et al. (1985); Dupont (1999); Angevine et al. (2003)) where decreasing the UHI intensity results
407 in a reduced PBLH difference between urban and non-urban areas.

408 **4. Implications for tracer transport and inverse modeling**

409 As expected, the different performance of each configuration is reflected in the tracer transport.
410 The daily cycle of the integrated footprint from the STILT model (Figure 10 (a)) reflects large
411 differences between configurations. The strongest daily cycle (largest amplitude) is the one for
412 BOUL+UCM while the weakest is for MYNN. During the night BOULC+UCM and YSU behave
413 similarly while QNSE and MYNN behave similarly to each other as well. On the other hand,
414 during the day the similarities are changed and YSU and MYNN show similar response while
415 QNSE resembles the BOUL+UCM values. Indeed, the differences with respect to MYNN (Figure
416 10 (b)) are between 10 to 50 % for YSU and BOUL+UCM during nighttime and about -30 %
417 for QNSE and BOUL+UCM during daytime. These differences are reflected in the CO₂ mole
418 fraction and thus in the bias as well. Figure 10 (c) shows that mean daily cycle for MYNN is the
419 least biased for all hours. During nighttime, MYNN shows a positive bias between 1 to 2 ppm
420 while the rest show much stronger biases with up to 6 ppm for BOUL+UCM. During daytime,
421 YSU still shows a positive bias about 1 to 2 ppm while MYNN fluctuates around 0 ppm. On the
422 other hand, BOUL+UCM and QNSE show negative biases between -1 and -4 ppm depending on
423 the hour.

424 In general, strong underestimation of PBLH during nights as shown by the configurations tested
425 in this work with the exception of MYNN, results in large accumulation of pollutants emitted from
426 local sources and thus strong nighttime positive bias. During daytime, the situation is different as
427 most models show a small relative PBLH bias. Nevertheless, the CO₂ daytime bias is non-zero and
428 different in direction depending on the configuration. The smaller daytime bias in most models

429 supports the typical practice in inverse modeling of only using afternoon hours. However, the
430 results shown here imply that MYNN has the potential of extending the inversion analysis to
431 nighttime as well due to the much smaller biases (and comparable to daytime) during this time of
432 the day.

433 The fact that positive UHI intensities are associated with deeper modeled PBLs over the urban
434 areas and that positive UHI are generally simulated by all models during nights implies that the
435 pollutant mixing during these hours in the urban areas is more active than in the rural counterpart.
436 In addition, it could also favor the development of urban centripetal circulations, as described in
437 Oke (1995), further impacting the pollutant advection. However, the fact that YSU is reproducing
438 in many occasions deeper nocturnal PBLHs over the non-urban areas would imply a more active
439 mixing outside of the city and the inhibition of the urban centripetal circulation.

440 In addition, inverse modeling based on the concept of footprints (observations' sensitivity to sur-
441 face fluxes) relies on Lagrangian Particle Dispersion Models (LPDMs) driven by meteorological
442 fields as those generated in this work. The footprints depend mostly on the advection of the par-
443 ticles (driven by the wind field), the turbulent mixing (driven by the turbulent velocity variances)
444 and the planetary boundary layer height (PBLH). Deeper modeled PBLHs than observed would
445 result in artificial dilution of the footprints and, therefore, source term overestimation. In addition,
446 some Lagrangian models parametrize the turbulent velocities as a function of the heat flux at the
447 surface. The fact that PBLH errors are mostly positively correlated with sensible heat flux errors
448 implies that an overestimation of the heat flux will cause an overestimation of the turbulent mixing
449 as well as PBLH, having a non-linear impact on the overall strength of the footprints.

450 **5. Conclusions**

451 We show that using ceilometers we were able to analyze the daily cycles of the PBLH and found
452 that most PBL schemes largely underestimate PBLH during nights. We also show that with these
453 measurements, correlations between PBLH errors and heat fluxes errors can be calculated and
454 serve to identify models that do not follow the proper trend. These results could not be obtained
455 using operational radiosondes as they are very limited in time (only twice a day).

456 We find the BEP urban canopy model did not improve the model performances in general and it
457 had an adverse impact on PBLH and sensible heat flux as compared to measurements. The UCM
458 partially corrected the Boulac nocturnal positive wind speed bias and negative PBLH bias at the
459 cost of increasing the negative bias as well as increasing the positive PBLH bias during daytime.
460 In addition, the UHI and ratio urban-rural of Bowen ratio did not compare well with the rest of
461 configurations or previously published results.

462 We find that modeled PBLHs are typically biased low during nighttime for most of the config-
463 urations with the exception of those using the MYNN parametrization. In addition, we find that
464 PBLH errors are mostly positively correlated with sensible heat flux errors, and that modeled pos-
465 itive UHI intensities were associated with deeper modeled PBLs over the urban areas. Overall, the
466 configurations using MYNN scheme performed the best, reproducing the PBLH reasonably well
467 during all hours.

468 We show that strong underestimation of PBLH during nights results in large accumulation of
469 pollutants emitted from local sources and thus strong nighttime positive CO₂ bias. However,
470 MYNN results suggest that, given the low night-time biases for this model, which are similar in
471 magnitude to the daytime biases, an inversion analysis may be extended into nighttime hours.

472 Last, we find that while most of the configurations performed as expected on reproducing the ur-
473 ban heat island effect, noticeable differences remain that may have an impact on weather and tracer
474 dispersion simulations in urban and regional studies. Further research is needed and experimental
475 intensive campaigns must be carried out to address these issues and differences as well as to better
476 understand the differences between PBL schemes during other seasons for the Washington, DC, /
477 Baltimore, MD, area.

478 *Acknowledgments.* We acknowledge the NWS Sterling Field Support Center (SFSC) and the
479 Howard University Beltsville Research center for providing the ceilometer data used in this work.
480 Funding was provided by the NIST Greenhouse Gas Measurements program.

481 Certain commercial equipment, instruments, or materials are identified in this paper in order
482 to specify the experimental procedure adequately. Such identification is not intended to imply
483 recommendation or endorsement by the National Institute of Standards and Technology, nor is it
484 intended to imply that the materials or equipment identified are necessarily the best available for
485 the purpose.

486 **References**

487 Angevine, W. M., L. Eddington, K. Durkee, C. Fairall, L. Bianco, and J. Brioude, 2012: Meteo-
488 rological model evaluation for CalNex 2010. *Monthly Weather Review*, **140** (12), 3885–3906,
489 doi:10.1175/mwr-d-12-00042.1, URL <https://doi.org/10.1175%2Fmwr-d-12-00042.1>.

490 Angevine, W. M., J. Olson, J. Kenyon, W. I. Gustafson, S. Endo, K. Suselj, and D. D. Turner, 2018:
491 Shallow cumulus in WRF parameterizations evaluated against LASSO large-eddy simulations.
492 *Monthly Weather Review*, **146** (12), 4303–4322, doi:10.1175/mwr-d-18-0115.1, URL [https://](https://doi.org/10.1175%2Fmwr-d-18-0115.1)
493 doi.org/10.1175%2Fmwr-d-18-0115.1.

- 494 Angevine, W. M., A. B. White, C. J. Senff, M. Trainer, R. M. Banta, and M. A. Ayoub,
495 2003: Urban-rural contrasts in mixing height and cloudiness over nashville in 1999. *Journal*
496 *of Geophysical Research: Atmospheres*, **108 (D3)**, n/a–n/a, doi:10.1029/2001jd001061, URL
497 <https://doi.org/10.1029%2F2001jd001061>.
- 498 Atkinson, D., B. Demoz, M. Hicks, and K. Vermeesch, 2017: Investigate and validate the effective-
499 ness of the vaisala cl31 ceilometer algorithm at selected sites across the u.s. for the automated
500 surface observing system (asos) program product improvement (phase 3). Tech. rep., National
501 Oceanic and Atmospheric Administration. URL <https://vlab.ncep.noaa.gov/group/cl31-project>.
- 502 Basara, J. B., P. K. Hall, A. J. Schroeder, B. G. Illston, and K. L. Nemunaitis, 2008: Diurnal
503 cycle of the oklahoma city urban heat island. *Journal of Geophysical Research*, **113 (D20)**,
504 doi:10.1029/2008jd010311, URL <https://doi.org/10.1029%2F2008jd010311>.
- 505 Benjamin, S. G., and Coauthors, 2016: A north american hourly assimilation and model fore-
506 cast cycle: The rapid refresh. *Monthly Weather Review*, **144 (4)**, 1669–1694, doi:10.1175/
507 mwr-d-15-0242.1, URL <https://doi.org/10.1175%2Fmwr-d-15-0242.1>.
- 508 Blaylock, B. K., J. D. Horel, and E. T. Crosman, 2017: Impact of lake breezes on summer ozone
509 concentrations in the salt lake valley. *Journal of Applied Meteorology and Climatology*, **56 (2)**,
510 353–370, doi:10.1175/jamc-d-16-0216.1, URL <https://doi.org/10.1175%2Fjamc-d-16-0216.1>.
- 511 Bougeault, P., and P. Lacarrere, 1989: Parameterization of orography-induced turbulence in a
512 mesobeta–scale model. *Monthly Weather Review*, **117 (8)**, 1872–1890.
- 513 Chen, F., and J. Dudhia, 2001: Coupling an advanced land surfacehydrology model with the penn
514 statencar mm5 modeling system. part i: Model implementation and sensitivity. *Monthly Weather*
515 *Review*, **129 (4)**, 569–585, doi:10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.

516 Compton, J. C., R. Delgado, T. A. Berkoff, and R. M. Hoff, 2013: Determination of planetary
517 boundary layer height on short spatial and temporal scales: A demonstration of the covariance
518 wavelet transform in ground-based wind profiler and lidar measurements. *Journal of Atmo-*
519 *spheric and Oceanic Technology*, **30** (7), 1566–1575, doi:10.1175/jtech-d-12-00116.1, URL
520 <https://doi.org/10.1175%2Fjtech-d-12-00116.1>.

521 Davis, K. J., N. Gamage, C. R. Hagelberg, C. Kiemle, D. H. Lenschow, and P. P.
522 Sullivan, 2000: An objective method for deriving atmospheric structure from airborne
523 lidar observations. *Journal of Atmospheric and Oceanic Technology*, **17** (11), 1455–
524 1468, doi:10.1175/1520-0426(2000)017<1455:AOMFDA>2.0.CO;2, URL [https://doi.org/10.](https://doi.org/10.1175/1520-0426(2000)017<1455:AOMFDA>2.0.CO;2)
525 [1175/1520-0426\(2000\)017<1455:AOMFDA>2.0.CO;2](https://doi.org/10.1175/1520-0426(2000)017<1455:AOMFDA>2.0.CO;2).

526 Díaz-Isaac, L. I., T. Lauvaux, and K. J. Davis, 2018: Impact of physical parameterizations and
527 initial conditions on simulated atmospheric transport and CO₂ mole fractions in the US midwest.
528 *Atmospheric Chemistry and Physics*, **18** (20), 14 813–14 835, doi:10.5194/acp-18-14813-2018,
529 URL <https://doi.org/10.5194%2FACP-18-14813-2018>.

530 Dupont, E., 1999: Comparison between the atmospheric boundary layer in paris and its rural sub-
531 urbs during the ECLAP experiment. *Atmospheric Environment*, **33** (6), 979–994, doi:10.1016/
532 [s1352-2310\(98\)00216-7](https://doi.org/10.1016%2Fs1352-2310(98)00216-7), URL <https://doi.org/10.1016%2Fs1352-2310%2898%2900216-7>.

533 Feng, S., and Coauthors, 2016: Los angeles megacity: a high-resolution land–atmosphere mod-
534 elling system for urban CO₂ emissions. *Atmospheric Chemistry and Physics*, **16** (14), 9019–
535 9045, doi:10.5194/acp-16-9019-2016, URL <https://doi.org/10.5194%2FACP-16-9019-2016>.

536 Gately, C. K., and L. R. Hutyrá, 2017: Large uncertainties in urban-scale carbon emis-
537 sions. *Journal of Geophysical Research: Atmospheres*, **122** (20), 11,242–11,260, doi:10.1002/
538 [2017jd027359](https://doi.org/10.1002%2F2017jd027359), URL <https://doi.org/10.1002%2F2017jd027359>.

539 Gerbig, C., J. C. Lin, S. C. Wofsy, B. C. Daube, A. E. Andrews, B. B. Stephens, P. S. Bakwin, and
540 C. A. Grainger, 2003: Toward constraining regional-scale fluxes of CO₂ with atmospheric obser-
541 vations over a continent: 2. analysis of COBRA data using a receptor-oriented framework. *Jour-
542 nal of Geophysical Research: Atmospheres*, **108** (D24), n/a–n/a, doi:10.1029/2003jd003770,
543 URL <https://doi.org/10.1029%2F2003jd003770>.

544 Godowitch, J. M., J. K. S. Ching, and J. F. Clarke, 1985: Evolution of the nocturnal inversion
545 layer at an urban and nonurban location. *Journal of Climate and Applied Meteorology*, **24** (8),
546 791–804, doi:10.1175/1520-0450(1985)024<0791:EOTNIL>2.0.CO;2.

547 Hegarty, J. D., and Coauthors, 2018: Analysis of the planetary boundary layer height dur-
548 ing DISCOVER-AQ baltimore–washington, d.c., with lidar and high-resolution WRF mod-
549 eling. *Journal of Applied Meteorology and Climatology*, **57** (11), 2679–2696, doi:10.1175/
550 jamc-d-18-0014.1, URL <https://doi.org/10.1175%2Fjamc-d-18-0014.1>.

551 Hicks, M., B. Demoz, K. Vermeesch, and D. Atkinson, 2019: Intercomparison of mixing layer
552 heights from the national weather service ceilometer test sites and collocated radiosondes. *Jour-
553 nal of Atmospheric and Oceanic Technology*, **36** (1), 129–137, doi:10.1175/jtech-d-18-0058.1,
554 URL <https://doi.org/10.1175%2Fjtech-d-18-0058.1>.

555 Hicks, M., R. Sakai, and E. Joseph, 2015: The evaluation of a new method to detect mix-
556 ing layer heights using lidar observations. *Journal of Atmospheric and Oceanic Technol-
557 ogy*, **32** (11), 2041–2051, doi:10.1175/jtech-d-14-00103.1, URL [https://doi.org/10.1175%
558 2Fjtech-d-14-00103.1](https://doi.org/10.1175%2Fjtech-d-14-00103.1).

559 Hong, S.-Y., 2010: A new stable boundary-layer mixing scheme and its impact on the simulated
560 east asian summer monsoon. *Quarterly Journal of the Royal Meteorological Society*, **136** (651),

561 1481–1496, doi:10.1002/qj.665, URL [https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.](https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.665)
562 665.

563 Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit
564 treatment of entrainment processes. *Monthly weather review*, **134** (9), 2318–2341.

565 Ito, J., H. Niino, M. Nakanishi, and C.-H. Moeng, 2015: An extension of the Mellor–Yamada
566 model to the terra incognita zone for dry convective mixed layers in the free convection regime.
567 *Boundary-Layer Meteorology*, **157** (1), 23–43, doi:10.1007/s10546-015-0045-5, URL <https://doi.org/10.1007/s10546-015-0045-5>.
568

569 Kain, J. S., 2004: The Kain–Fritsch convective parameterization: an update. *Journal of Applied*
570 *Meteorology*, **43** (1), 170–181.

571 Karion, A., and Coauthors, 2020: Greenhouse gas observations from the northeast corridor tower
572 network. *Earth System Science Data*, **12** (1), 699–717, doi:10.5194/essd-12-699-2020, URL
573 <https://doi.org/10.5194/essd-12-699-2020>.

574 Kretschmer, R., C. Gerbig, U. Karstens, G. Biavati, A. Vermeulen, F. Vogel, S. Hammer,
575 and K. U. Totsche, 2014: Impact of optimized mixing heights on simulated regional at-
576 mospheric transport of CO₂. *Atmospheric Chemistry and Physics*, **14** (14), 7149–7172, doi:
577 10.5194/acp-14-7149-2014, URL <https://doi.org/10.5194/acp-14-7149-2014>.

578 Kretschmer, R., C. Gerbig, U. Karstens, and F.-T. Koch, 2012: Error characterization of CO₂
579 vertical mixing in the atmospheric transport model WRF-VPRM. *Atmospheric Chemistry and*
580 *Physics*, **12** (5), 2441–2458, doi:10.5194/acp-12-2441-2012, URL [https://doi.org/10.5194/](https://doi.org/10.5194/acp-12-2441-2012)
581 [2/acp-12-2441-2012](https://doi.org/10.5194/acp-12-2441-2012).

- 582 Lauvaux, T., and Coauthors, 2016: High-resolution atmospheric inversion of urban CO₂ emis-
583 sions during the dormant season of the indianapolis flux experiment (INFLUX). *Journal of*
584 *Geophysical Research: Atmospheres*, **121** (10), 5213–5236, doi:10.1002/2015jd024473, URL
585 <https://doi.org/10.1002%2F2015jd024473>.
- 586 Lian, J., L. Wu, F.-M. Bréon, G. Broquet, R. Vautard, T. S. Zaccheo, J. Dobler, and P. Ciais, 2018:
587 Evaluation of the WRF-UCM mesoscale model and ECMWF global operational forecasts over
588 the paris region in the prospect of tracer atmospheric transport modeling. *Elem Sci Anth*, **6** (1),
589 64, doi:10.1525/elementa.319, URL <https://doi.org/10.1525%2Felementa.319>.
- 590 Lin, J. C., 2003: A near-field tool for simulating the upstream influence of atmospheric obser-
591 vations: The stochastic time-inverted lagrangian transport (STILT) model. *Journal of Geo-*
592 *physical Research*, **108** (D16), ACH 2–1–ACH 2–17, doi:10.1029/2002jd003161, URL <https://doi.org/10.1029%2F2002jd003161>.
593 [//doi.org/10.1029%2F2002jd003161](https://doi.org/10.1029%2F2002jd003161).
- 594 Lopez-Coto, I., S. Ghosh, K. Prasad, and J. Whetstone, 2017a: Tower-based greenhouse
595 gas measurement network design—the national institute of standards and technology north
596 east corridor testbed. *Advances in Atmospheric Sciences*, **34** (9), 1095–1105, doi:10.1007/
597 s00376-017-6094-6, URL <https://doi.org/10.1007%2Fs00376-017-6094-6>.
- 598 Lopez-Coto, I., K. Prasad, and J. R. Whetstone, 2017b: Carbon dioxide biogenic vs anthropogenic
599 sectoral contribution for the indianapolis flux experiment (INFLUX). Tech. Rep. NIST-SP 1237,
600 National Institute of Standards and Technology. doi:10.6028/nist.sp.1237, URL <https://doi.org/10.6028%2Fnist.sp.1237>.
601 [10.6028%2Fnist.sp.1237](https://doi.org/10.6028%2Fnist.sp.1237).
- 602 Martilli, A., A. Clappier, and M. W. Rotach, 2002: An urban surface exchange parameterisa-
603 tion for mesoscale models. *Boundary-layer meteorology*, **104** (2), 261–304, doi:10.1023/A:
604 1016099921195.

605 McMillen, R. T., 1988: An eddy correlation technique with extended applicability to non-simple
606 terrain. *Boundary-Layer Meteorology*, **43 (3)**, 231–245, doi:10.1007/BF00128405.

607 Mesinger, F., and Coauthors, 2006: North american regional reanalysis. *Bulletin of the American*
608 *Meteorological Society*, **87 (3)**, 343–360, doi:10.1175/bams-87-3-343, URL [https://doi.org/10.](https://doi.org/10.1175%2Fbams-87-3-343)
609 [1175%2Fbams-87-3-343](https://doi.org/10.1175%2Fbams-87-3-343).

610 Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer
611 for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Jour-*
612 *nal of Geophysical Research: Atmospheres*, **102 (D14)**, 16 663–16 682, doi:10.1029/97jd00237,
613 URL <https://doi.org/10.1029%2F97jd00237>.

614 Nakanishi, M., and H. Niino, 2004: An improved mellor–yamada level-3 model with condensation
615 physics: Its design and verification. *Boundary-layer meteorology*, **112 (1)**, 1–31.

616 Nakanishi, M., and H. Niino, 2006: An improved mellor–yamada level-3 model: Its numeri-
617 cal stability and application to a regional prediction of advection fog. *Boundary-Layer Me-*
618 *teorology*, **119 (2)**, 397–407, doi:10.1007/s10546-005-9030-8, URL [https://doi.org/10.1007%](https://doi.org/10.1007%2Fs10546-005-9030-8)
619 [2Fs10546-005-9030-8](https://doi.org/10.1007%2Fs10546-005-9030-8).

620 Nehr Korn, T., J. Eluszkiewicz, S. C. Wofsy, J. C. Lin, C. Gerbig, M. Longo, and S. Freitas,
621 2010: Coupled weather research and forecasting–stochastic time-inverted lagrangian trans-
622 port (WRF–STILT) model. *Meteorology and Atmospheric Physics*, **107 (1-2)**, 51–64, doi:
623 [10.1007/s00703-010-0068-x](https://doi.org/10.1007/s00703-010-0068-x), URL <https://doi.org/10.1007%2Fs00703-010-0068-x>.

624 Nisbet, E., and R. Weiss, 2010: Top-down versus bottom-up. *Science*, **328 (5983)**, 1241–1243,
625 doi:10.1126/science.1189936, URL <https://doi.org/10.1126%2Fscience.1189936>.

- 626 Oke, T. R., 1982: The energetic basis of the urban heat island. *Quarterly Journal of the Royal*
627 *Meteorological Society*, **108 (455)**, 1–24, doi:10.1002/qj.49710845502, URL [https://doi.org/10.](https://doi.org/10.1002%2Fqj.49710845502)
628 [1002%2Fqj.49710845502](https://doi.org/10.1002%2Fqj.49710845502).
- 629 Oke, T. R., 1995: The heat island of the urban boundary layer: Characteristics, causes and effects.
630 *Wind Climate in Cities*, Springer Netherlands, 81–107, doi:10.1007/978-94-017-3686-2_5,
631 URL https://doi.org/10.1007%2F978-94-017-3686-2_5.
- 632 Olson, J. B., J. S. Kenyon, W. A. Angevine, J. M. Brown, M. Pagowski, and K. Suelj, 2019: A
633 Description of the MYNN-EDMF Scheme and the Coupling to Other Components in WR-
634 FARW. Tech. Rep. OAR GSD-61, National Oceanic and Atmospheric Administration. doi:
635 [10.25923/n9wm-be49](https://doi.org/10.25923/n9wm-be49).
- 636 Salamanca, F., A. Martilli, M. Tewari, and F. Chen, 2011a: A study of the urban bound-
637 ary layer using different urban parameterizations and high-resolution urban canopy parame-
638 ters with WRF. *Journal of Applied Meteorology and Climatology*, **50 (5)**, 1107–1128, doi:
639 [10.1175/2010jamc2538.1](https://doi.org/10.1175/2010jamc2538.1), URL <https://doi.org/10.1175%2F2010jamc2538.1>.
- 640 Salamanca, F., A. Martilli, and C. Yage, 2011b: A numerical study of the urban heat island over
641 madrid during the DESIREX (2008) campaign with WRF and an evaluation of simple mitigation
642 strategies. *International Journal of Climatology*, **32 (15)**, 2372–2386, doi:10.1002/joc.3398,
643 URL <https://doi.org/10.1002%2Fjoc.3398>.
- 644 Sarmiento, D. P., K. J. Davis, A. Deng, T. Lauvaux, A. Brewer, and M. Hardesty, 2017: A
645 comprehensive assessment of land surface-atmosphere interactions in a WRF/urban model-
646 ing system for indianapolis, IN. *Elem Sci Anth*, **5**, doi:10.1525/elementa.132, URL <https://doi.org/10.1525%2Felementa.132>.
647 [//doi.org/10.1525%2Felementa.132](https://doi.org/10.1525%2Felementa.132).

648 Seidel, D. J., Y. Zhang, A. Beljaars, J.-C. Golaz, A. R. Jacobson, and B. Medeiros, 2012: Clima-
649 tology of the planetary boundary layer over the continental united states and europe. *Journal*
650 *of Geophysical Research: Atmospheres*, **117 (D17)**, n/a–n/a, doi:10.1029/2012jd018143, URL
651 <https://doi.org/10.1029%2F2012jd018143>.

652 Shin, H. H., and S.-Y. Hong, 2011: Intercomparison of planetary boundary-layer parametriza-
653 tions in the WRF model for a single day from CASES-99. *Boundary-Layer Meteo-*
654 *rology*, **139 (2)**, 261–281, doi:10.1007/s10546-010-9583-z, URL [https://doi.org/10.1007%](https://doi.org/10.1007%2Fs10546-010-9583-z)
655 [2Fs10546-010-9583-z](https://doi.org/10.1007%2Fs10546-010-9583-z).

656 Skamarock, W. C., and Coauthors, 2008: A description of the advanced research wrf version 3.
657 Tech. Rep. NCAR/TN-475+STR, National Center for Atmospheric Research.

658 Smith, A., N. Lott, and R. Vose, 2011: The integrated surface database: Recent developments and
659 partnerships. *Bulletin of the American Meteorological Society*, **92 (6)**, 704–708, doi:10.1175/
660 2011bams3015.1, URL <https://doi.org/10.1175%2F2011bams3015.1>.

661 Spangler, T. C., and R. A. Dirks, 1974: Meso-scale variations of the urban mixing height.
662 *Boundary-Layer Meteorology*, **6 (3-4)**, 423–441, doi:10.1007/bf02137677, URL [https://doi.org/](https://doi.org/10.1007%2Fbf02137677)
663 [10.1007%2Fbf02137677](https://doi.org/10.1007%2Fbf02137677).

664 Steyn, D. G., M. Baldi, and R. Hoff, 1999: The detection of mixed layer depth and entrainment
665 zone thickness from lidar backscatter profiles. *Journal of Atmospheric and Oceanic Technology*,
666 **16 (7)**, 953–959.

667 Stull, R. B., Ed., 1988: *An Introduction to Boundary Layer Meteorology*. Springer Netherlands,
668 doi:10.1007/978-94-009-3027-8, URL <https://doi.org/10.1007%2F978-94-009-3027-8>.

669 Sukoriansky, S., B. Galperin, and I. Staroselsky, 2005: A quasinormal scale elimination model
670 of turbulent flows with stable stratification. *Physics of fluids*, **17 (8)**, 085–107, doi:10.1063/1.
671 2009010.

672 Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter
673 precipitation using an improved bulk microphysics scheme. part ii: Implementation of a new
674 snow parameterization. *Monthly Weather Review*, **136 (12)**, 5095–5115.

675 Thompson, G., R. M. Rasmussen, and K. Manning, 2004: Explicit forecasts of winter precipita-
676 tion using an improved bulk microphysics scheme. part i: Description and sensitivity analysis.
677 *Monthly Weather Review*, **132 (2)**, 519–542.

678 Ware, J., E. A. Kort, P. DeCola, and R. Duren, 2016: Aerosol lidar observations of atmospheric
679 mixing in los angeles: Climatology and implications for greenhouse gas observations. *Journal of*
680 *Geophysical Research: Atmospheres*, **121 (16)**, 9862–9878, doi:10.1002/2016jd024953, URL
681 <https://doi.org/10.1002%2F2016jd024953>.

682 Webb, E. K., G. I. Pearman, and R. Leuning, 1980: Correction of flux measurements for density
683 effects due to heat and water vapour transfer. *Quarterly Journal of the Royal Meteorological*
684 *Society*, **106 (447)**, 85–100.

685 Wood, C. R., and Coauthors, 2013: An overview of the urban boundary layer atmosphere network
686 in helsinki. *Bulletin of the American Meteorological Society*, **94 (11)**, 1675–1690, doi:10.1175/
687 bams-d-12-00146.1, URL <https://doi.org/10.1175%2Fbams-d-12-00146.1>.

688 Zhang, D.-L., Y.-X. Shou, and R. R. Dickerson, 2009: Upstream urbanization exacerbates urban
689 heat island effects. *Geophysical Research Letters*, **36 (24)**, doi:10.1029/2009gl041082, URL
690 <https://doi.org/10.1029%2F2009gl041082>.

691 Zhang, D.-L., Y.-X. Shou, R. R. Dickerson, and F. Chen, 2011: Impact of upstream urban-
692 ization on the urban heat island effects along the washington–baltimore corridor. *Journal of*
693 *Applied Meteorology and Climatology*, **50** (10), 2012–2029, doi:10.1175/jamc-d-10-05008.1,
694 URL <https://doi.org/10.1175%2Fjamc-d-10-05008.1>.

695 **LIST OF TABLES**

696 **Table 1.** WRF model configurations 36

697 **Table 2.** ISD statistics mean bias (model - observations), standard deviation (SD) and
698 the interquartile range (IQR) of the differences. February 2016. 37

699 **Table 3.** Global statistics for the PBLH errors (model minus observed). February 2016. . . . 38

700 **Table 4.** Global statistics for the sensible (HFX) and latent (LE) heat fluxes errors
701 (model - observed) at Beltsville (HUBV) (units: $W\ m^{-2}$). February 2016. 39

TABLE 1. WRF model configurations

| Label | Version | PBL scheme | Surface Layer | IC/BC | Land use | Urban canopy model |
|-------------------------|---------|------------|---------------|-------|----------|--------------------|
| YSU | 3.8 | YSU | MOST | HRRR | USGS | — |
| YSU+NARR | 3.8 | YSU | MOST | NARR | USGS | — |
| MYNN | 3.8 | MYNN | MYNN | HRRR | USGS | — |
| MYNN _e | 3.9.1.1 | MYNN+edmf | MYNN | NARR | USGS | — |
| MYNN _e +nlcd | 3.9.1.1 | MYNN+edmf | MYNN | NARR | NLCD | — |
| BOUL | 3.8 | BouLac | MOST | HRRR | USGS | — |
| BOUL+UCM | 3.8 | BouLac | MOST | HRRR | USGS33 | BEP |
| QNSE | 3.8 | QNSE | QNSE | HRRR | USGS | — |

702 TABLE 2. ISD statistics mean bias (model - observations), standard deviation (SD) and the interquartile range
 703 (IQR) of the differences. February 2016.

| | | YSU | YSU+NARR | MYNN | MYNNe | MYNNe+nlcd | BOUL | BOUL+UCM | QNSE |
|-------------------|------|-------|----------|-------|-------|------------|-------|----------|-------|
| T (K) | Bias | 0.09 | -0.63 | -0.56 | -0.55 | -0.92 | 1.26 | 1.96 | -0.71 |
| | SD | 1.60 | 2.04 | 1.76 | 2.04 | 2.00 | 1.92 | 2.02 | 1.62 |
| | IQR | 1.63 | 2.16 | 1.84 | 2.16 | 2.33 | 2.03 | 2.30 | 1.68 |
| ws (m/s) | Bias | -0.45 | -0.53 | -0.88 | -0.60 | -1.12 | 0.69 | -0.70 | -0.91 |
| | SD | 1.70 | 1.81 | 1.96 | 1.73 | 1.72 | 2.00 | 2.16 | 1.70 |
| | IQR | 2.03 | 2.17 | 2.37 | 2.02 | 2.12 | 2.31 | 2.40 | 1.99 |
| wd ($^{\circ}$) | Bias | 1.04 | 0.99 | -1.42 | -0.07 | -3.27 | 5.79 | 6.43 | -5.03 |
| | SD | 36.64 | 42.11 | 47.12 | 41.63 | 43.99 | 36.62 | 37.53 | 40.03 |
| | IQR | 27.91 | 36.42 | 34.76 | 37.26 | 42.84 | 27.49 | 28.80 | 29.73 |

TABLE 3. Global statistics for the PBLH errors (model minus observed). February 2016.

| BOTH | YSU | YSU+NARR | MYNN | MYNNe | MYNNe+nlcd | BOUL | BOUL+UCM | QNSE |
|------------|-----|----------|------|-------|------------|------|----------|------|
| Co (%) | 56 | 55 | 60 | 57 | 59 | 52 | 55 | 52 |
| Mean (m) | -38 | -73 | -57 | -32 | 15 | -50 | 50 | 152 |
| SD (m) | 494 | 496 | 462 | 496 | 480 | 519 | 513 | 542 |
| Median (m) | -93 | -119 | -72 | -89 | -47 | -126 | -25 | 77 |
| IQR (m) | 484 | 502 | 460 | 462 | 479 | 516 | 571 | 630 |

| SFSC | YSU | YSU+NARR | MYNN | MYNNe | MYNNe+nlcd | BOUL | BOUL+UCM | QNSE |
|------------|-----|----------|------|-------|------------|------|----------|------|
| Co (%) | 50 | 51 | 58 | 51 | 56 | 46 | 48 | 47 |
| Mean (m) | -17 | -41 | -27 | -11 | 43 | -15 | 53 | 163 |
| SD (m) | 497 | 509 | 459 | 528 | 479 | 535 | 523 | 545 |
| Median (m) | -94 | -121 | -69 | -110 | -37 | -116 | -46 | 96 |
| IQR (m) | 579 | 541 | 469 | 530 | 519 | 623 | 655 | 692 |

| HUBV | YSU | YSU+NARR | MYNN | MYNNe | MYNNe+nlcd | BOUL | BOUL+UCM | QNSE |
|------------|-----|----------|------|-------|------------|------|----------|------|
| Co (%) | 63 | 60 | 62 | 63 | 61 | 58 | 61 | 56 |
| Mean (m) | -61 | -107 | -88 | -55 | -15 | -86 | 48 | 139 |
| SD (m) | 491 | 480 | 463 | 460 | 480 | 499 | 503 | 539 |
| Median (m) | -91 | -114 | -77 | -80 | -56 | -133 | -9 | 59 |
| IQR (m) | 396 | 459 | 440 | 416 | 433 | 411 | 481 | 554 |

704 TABLE 4. Global statistics for the sensible (HFX) and latent (LE) heat fluxes errors (model - observed) at
 705 Beltsville (HUBV) (units: W m^{-2}). February 2016.

| HFX | YSU | YSU+NARR | MYNN | MYNNe | MYNNe+nlcd | BOUL | BOUL+UCM | QNSE |
|--------|-----|----------|------|-------|------------|------|----------|------|
| Mean | -13 | -15 | -23 | -16 | 12 | -19 | 8 | -20 |
| SD | 90 | 86 | 96 | 89 | 96 | 94 | 97 | 92 |
| Median | -2 | -2 | -9 | -5 | 16 | -7 | 3 | -10 |
| IQR | 46 | 48 | 56 | 51 | 65 | 57 | 61 | 61 |

| LE | YSU | YSU+NARR | MYNN | MYNNe | MYNNe+nlcd | BOUL | BOUL+UCM | QNSE |
|--------|-----|----------|------|-------|------------|------|----------|------|
| Mean | 8 | 9 | 11 | 13 | -11 | 13 | 0.5 | 21 |
| SD | 58 | 59 | 61 | 61 | 58 | 58 | 57 | 65 |
| Median | 5 | 6 | 9 | 8 | 0.2 | 11 | 0.6 | 12 |
| IQR | 23 | 27 | 26 | 30 | 22 | 29 | 16 | 39 |

706 **LIST OF FIGURES**

707 **Fig. 1.** Innermost computational domain (1 km) showing the urban land use categories employed in
 708 the model: left, USGS and right, NLCD for the area of interest (Washington DC / Baltimore
 709 region) along with the location of the ISD surface stations and two ceilometers used for
 710 verification. 41

711 **Fig. 2.** Median daily cycle of the differences for (a) temperature, (b) wind speed, (c) wind direction,
 712 (d) planetary boundary layer heights, (e) sensible heat flux and (f) latent heat flux. Grey lines
 713 correspond to the rest of configurations not present in the legend. February 2016. 42

714 **Fig. 3.** Daily cycle for the observed Planetary Boundary Layer Heights (PBLH) at the two ceilome-
 715 ter locations for the month of February 2016. Red line is the median, blue bar represents the
 716 first and third quantile and the whisker bars are limited to 1.5 times the IQR. February 2016. 43

717 **Fig. 4.** Daily cycle of the relative differences for the PBLH, including both ceilometers, for the eight
 718 configurations tested. Red line is the median, blue bar represents the first and third quantile
 719 and the whisker bars are limited to 1.5 times the IQR. February 2016. 44

720 **Fig. 5.** Scatter plot of the mean daily cycle of PBLH differences vs the sensible heat flux differences
 721 at Beltsville (HUBV). February 2016. 45

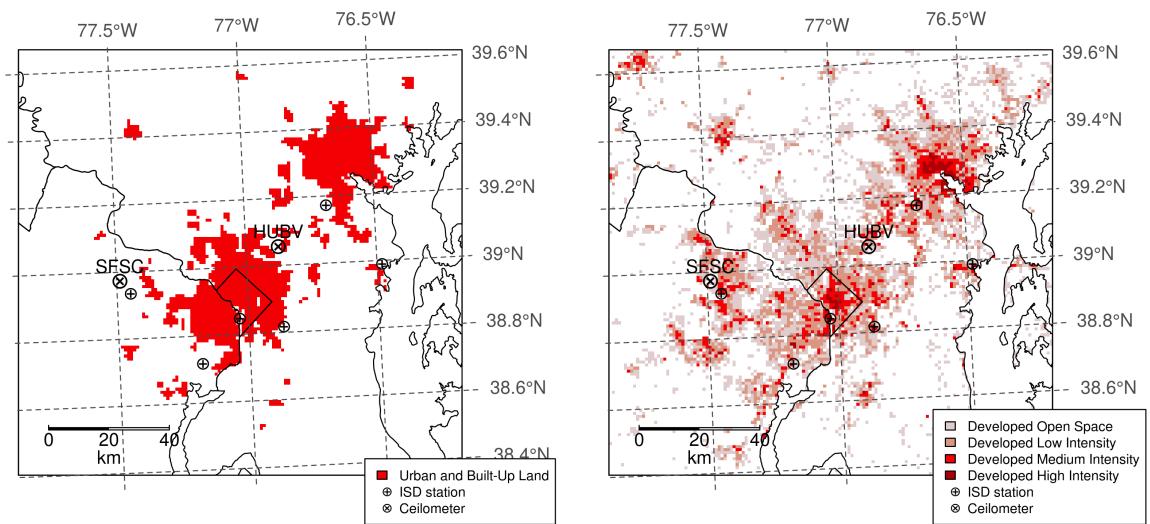
722 **Fig. 6.** Median daily cycle for the (a) urban heat island (UHI), (b) the area averaged Bowen ratio for
 723 urban areas over Bowen ratio for non-urban areas (β_{ua}/β_{nua}), (c) the area averaged sensible
 724 heat flux difference between urban and non-urban areas (ΔHFX) and (d) the area averaged
 725 latent heat flux difference between urban and non-urban areas (ΔLE). Grey lines correspond
 726 to the rest of configurations not present in the legend. Note that in (b) the black horizontal
 727 dashed line marks the value 1. February 2016. 46

728 **Fig. 7.** Daily cycle for the area averaged PBLH difference between urban and non-urban areas
 729 ($\Delta PBLH$). Red line is the median, blue bar represents the first and third quantile and the
 730 whisker bars are limited to 1.5 times the IQR. February 2016. 47

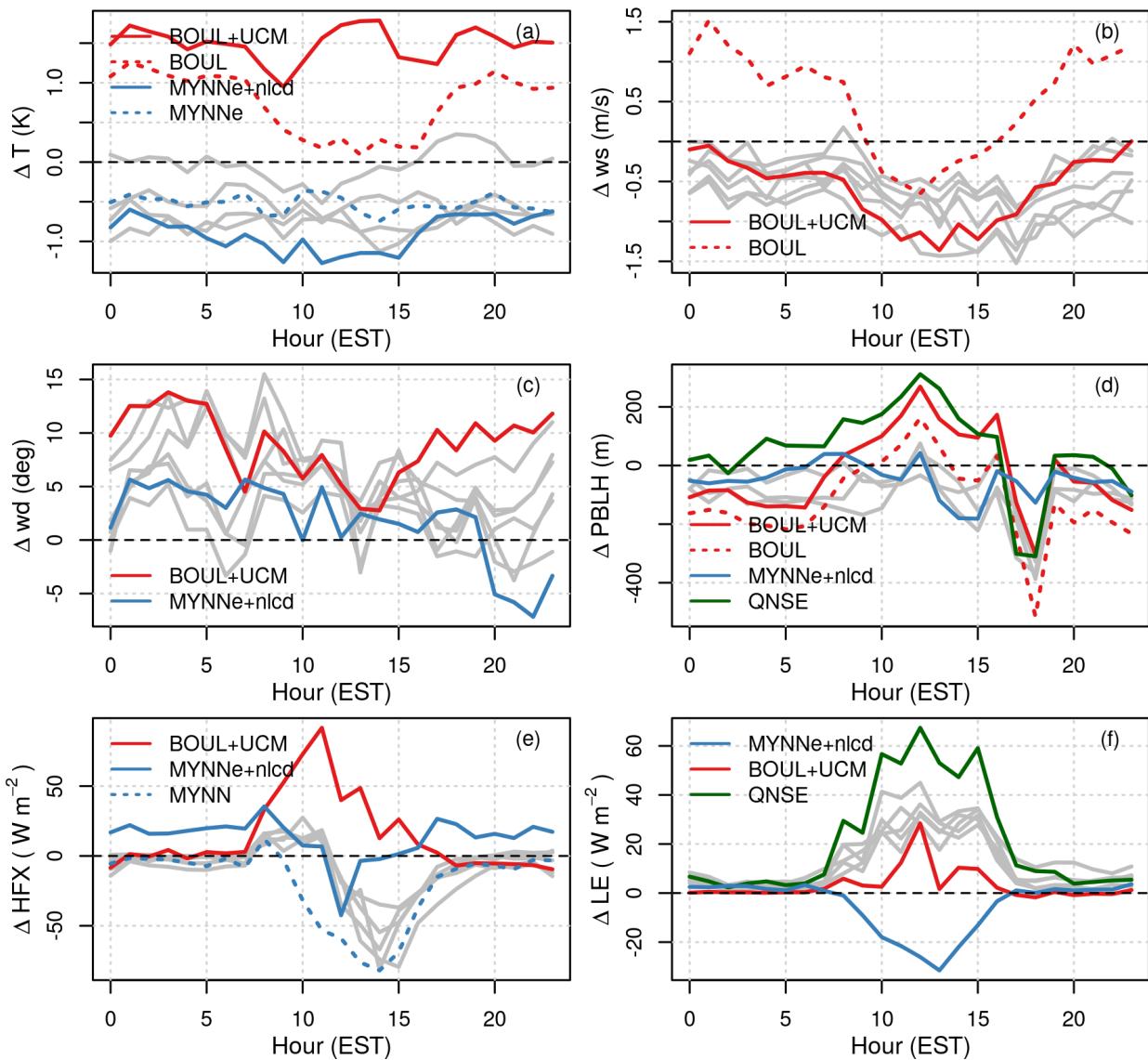
731 **Fig. 8.** Scatter plot and linear model fit of the relationship between the UHI and the $\Delta PBLH$. Febru-
 732 ary 2016. 48

733 **Fig. 9.** Median daily cycle of the UHI and the PBLH difference between urban and non-urban areas
 734 ($\Delta PBLH$). The marked point represents the time 19 EST while the line connects the points
 735 as the time increases, thus ending at 18 EST. February 2016. 49

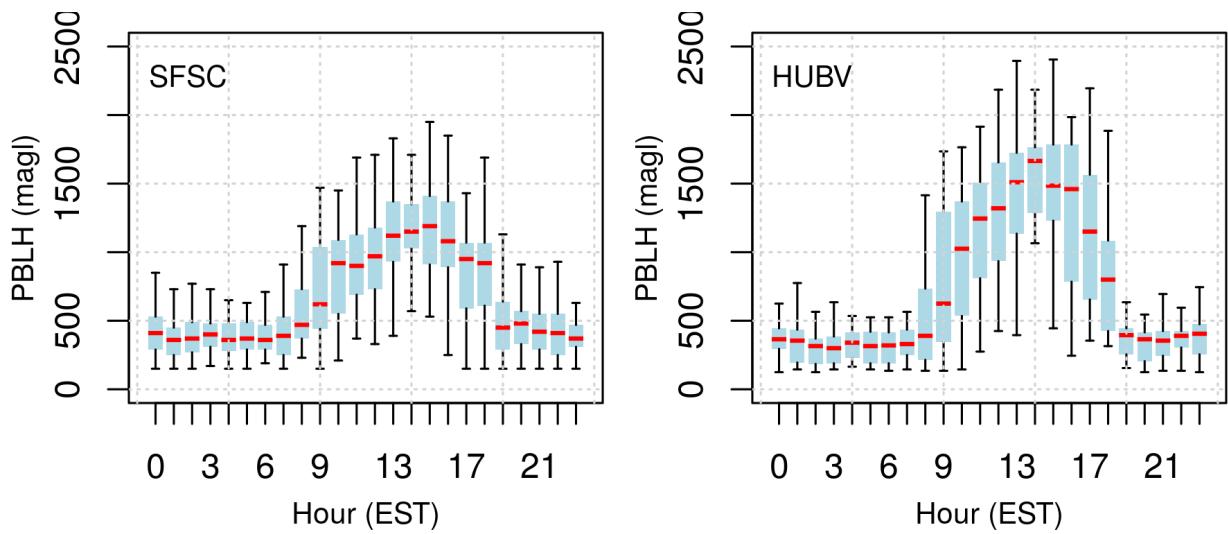
736 **Fig. 10.** Daily cycle of the a) Integrated footprint (ppm), b) relative difference to MYNN (%),c) CO
 737 ₂ mean bias (ppm) and, d) CO ₂ median bias (ppm) for the two towers and 4 configurations.
 738 February 2016. 50



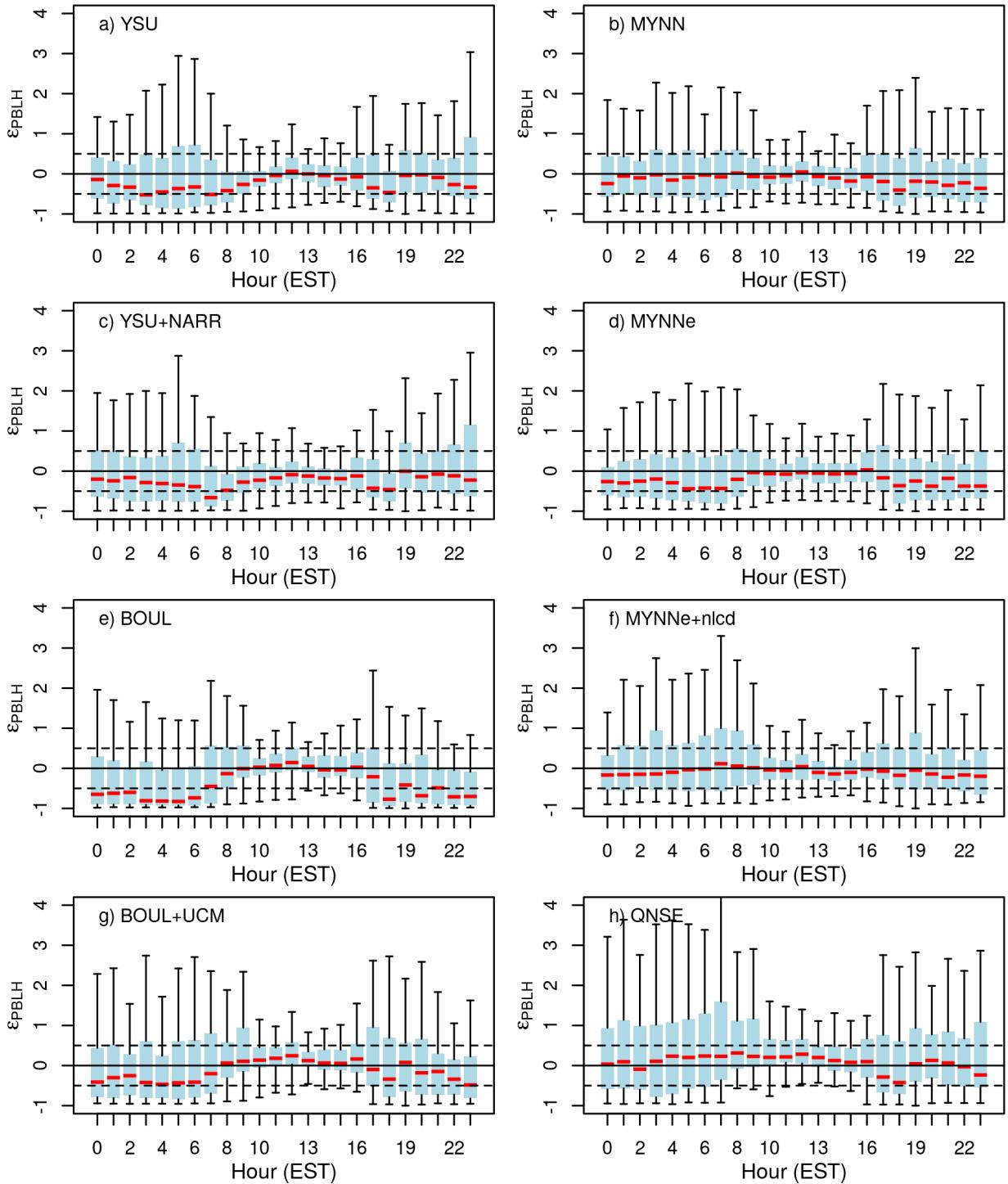
739 FIG. 1. Innermost computational domain (1 km) showing the urban land use categories employed in the
 740 model: left, USGS and right, NLCD for the area of interest (Washington DC / Baltimore region) along with the
 741 location of the ISD surface stations and two ceilometers used for verification.



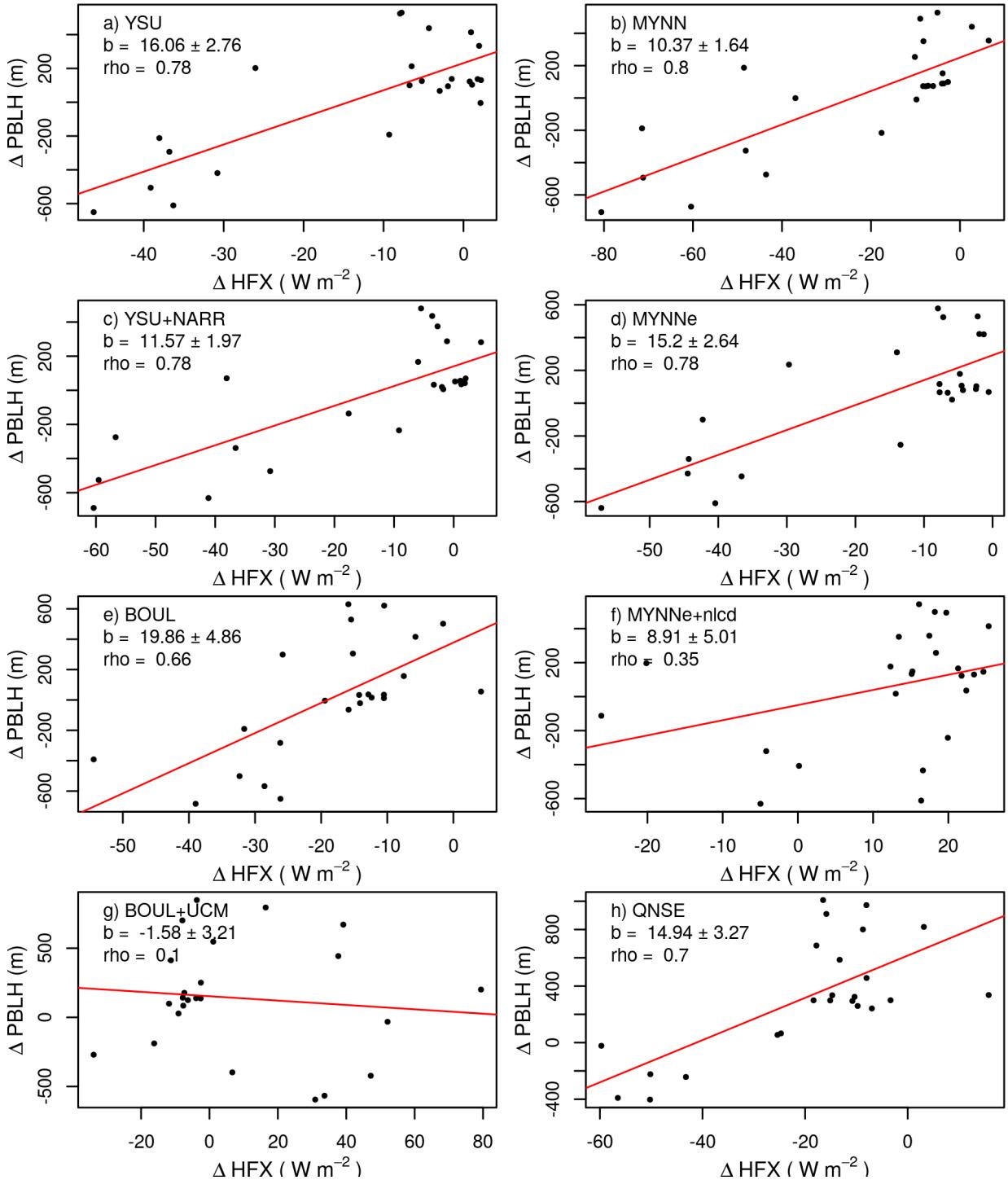
742 FIG. 2. Median daily cycle of the differences for (a) temperature, (b) wind speed, (c) wind direction, (d)
 743 planetary boundary layer heights, (e) sensible heat flux and (f) latent heat flux. Grey lines correspond to the rest
 744 of configurations not present in the legend. February 2016.



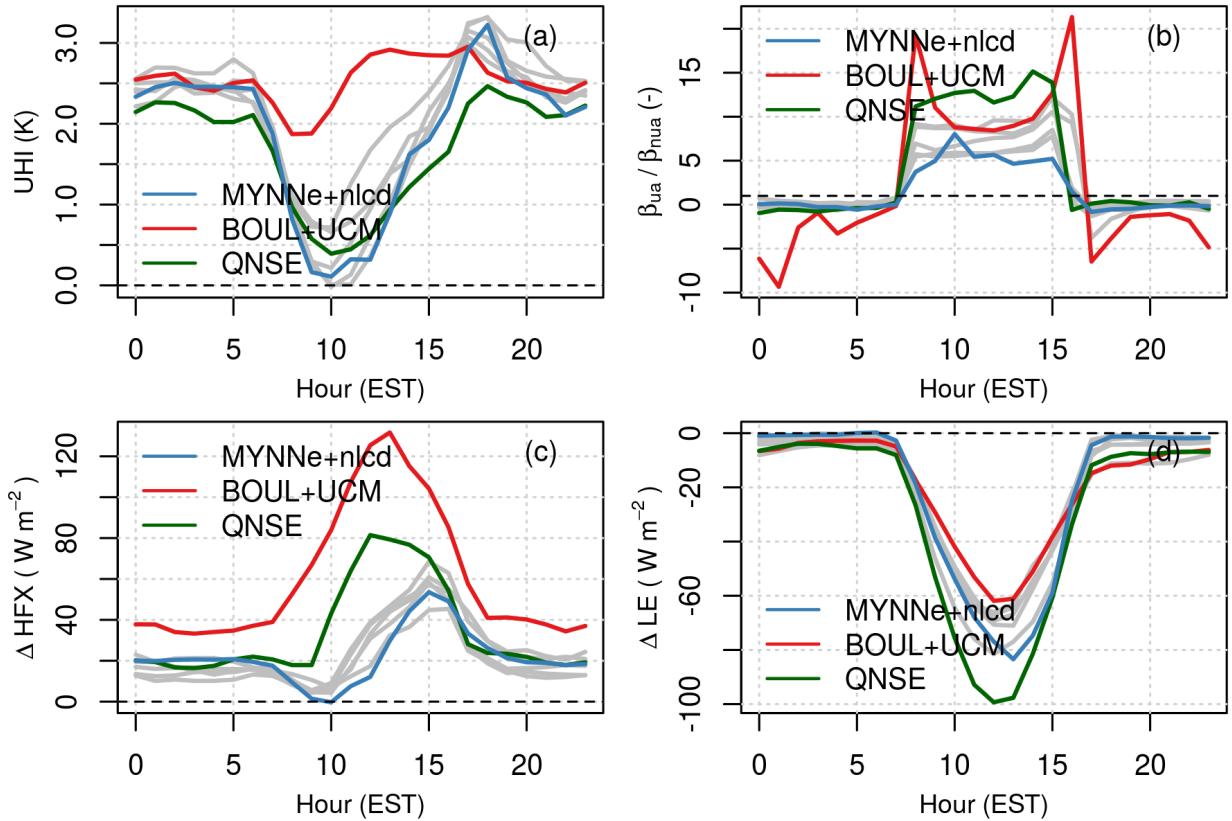
745 FIG. 3. Daily cycle for the observed Planetary Boundary Layer Heights (PBLH) at the two ceilometer loca-
 746 tions for the month of February 2016. Red line is the median, blue bar represents the first and third quartile and
 747 the whisker bars are limited to 1.5 times the IQR. February 2016.



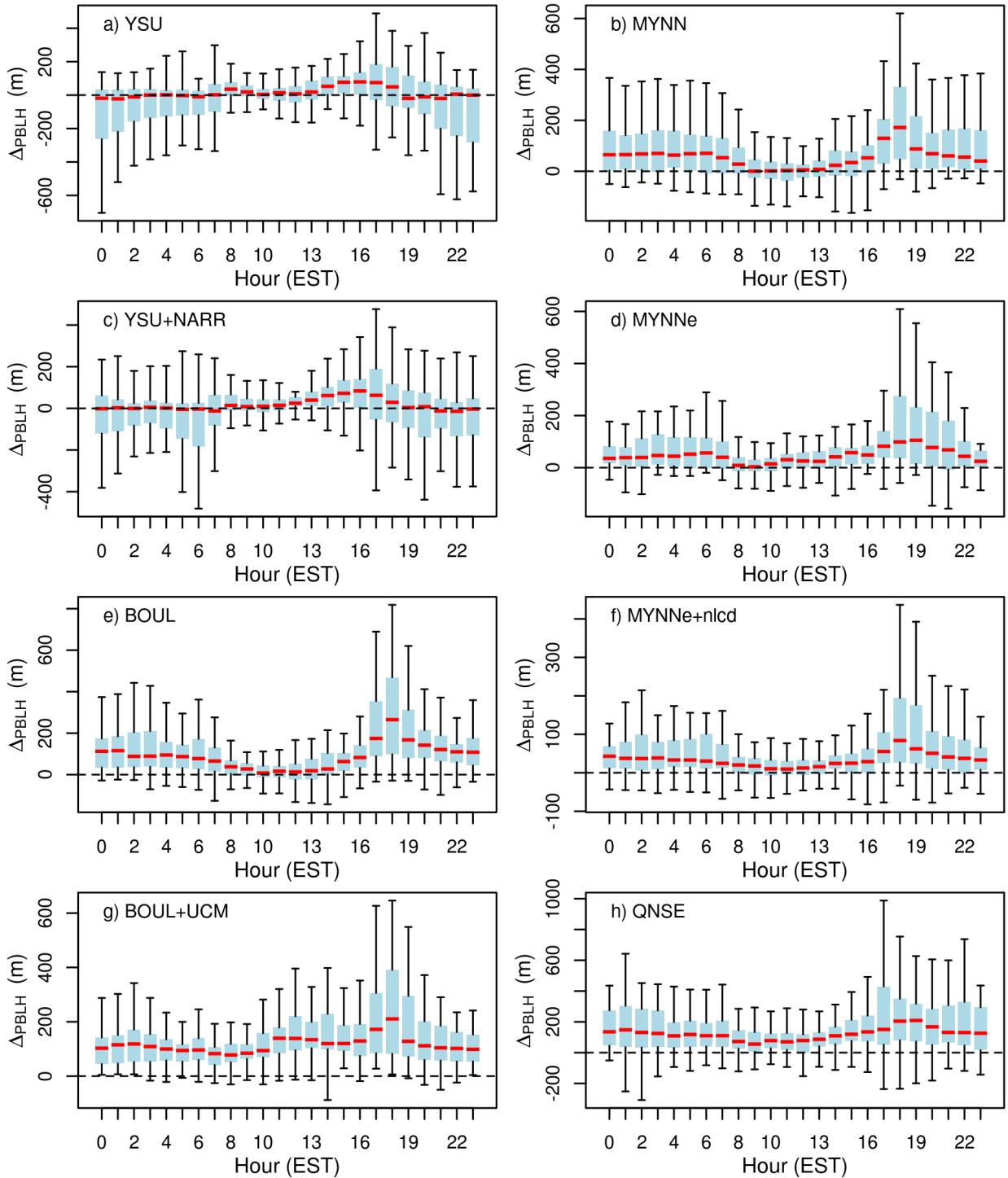
748 FIG. 4. Daily cycle of the relative differences for the PBLH, including both ceilometers, for the eight config-
 749 urations tested. Red line is the median, blue bar represents the first and third quantile and the whisker bars are
 750 limited to 1.5 times the IQR. February 2016.



751 FIG. 5. Scatter plot of the mean daily cycle of PBLH differences vs the sensible heat flux differences at
 752 Beltsville (HUBV). February 2016.



753 FIG. 6. Median daily cycle for the (a) urban heat island (UHI), (b) the area averaged Bowen ratio for urban
 754 areas over Bowen ratio for non-urban areas (β_{ua}/β_{nua}), (c) the area averaged sensible heat flux difference be-
 755 tween urban and non-urban areas (ΔHFX) and (d) the area averaged latent heat flux difference between urban
 756 and non-urban areas (ΔLE). Grey lines correspond to the rest of configurations not present in the legend. Note
 757 that in (b) the black horizontal dashed line marks the value 1. February 2016.



758 FIG. 7. Daily cycle for the area averaged PBLH difference between urban and non-urban areas ($\Delta PBLH$). Red
 759 line is the median, blue bar represents the first and third quantile and the whisker bars are limited to 1.5 times
 760 the IQR. February 2016.

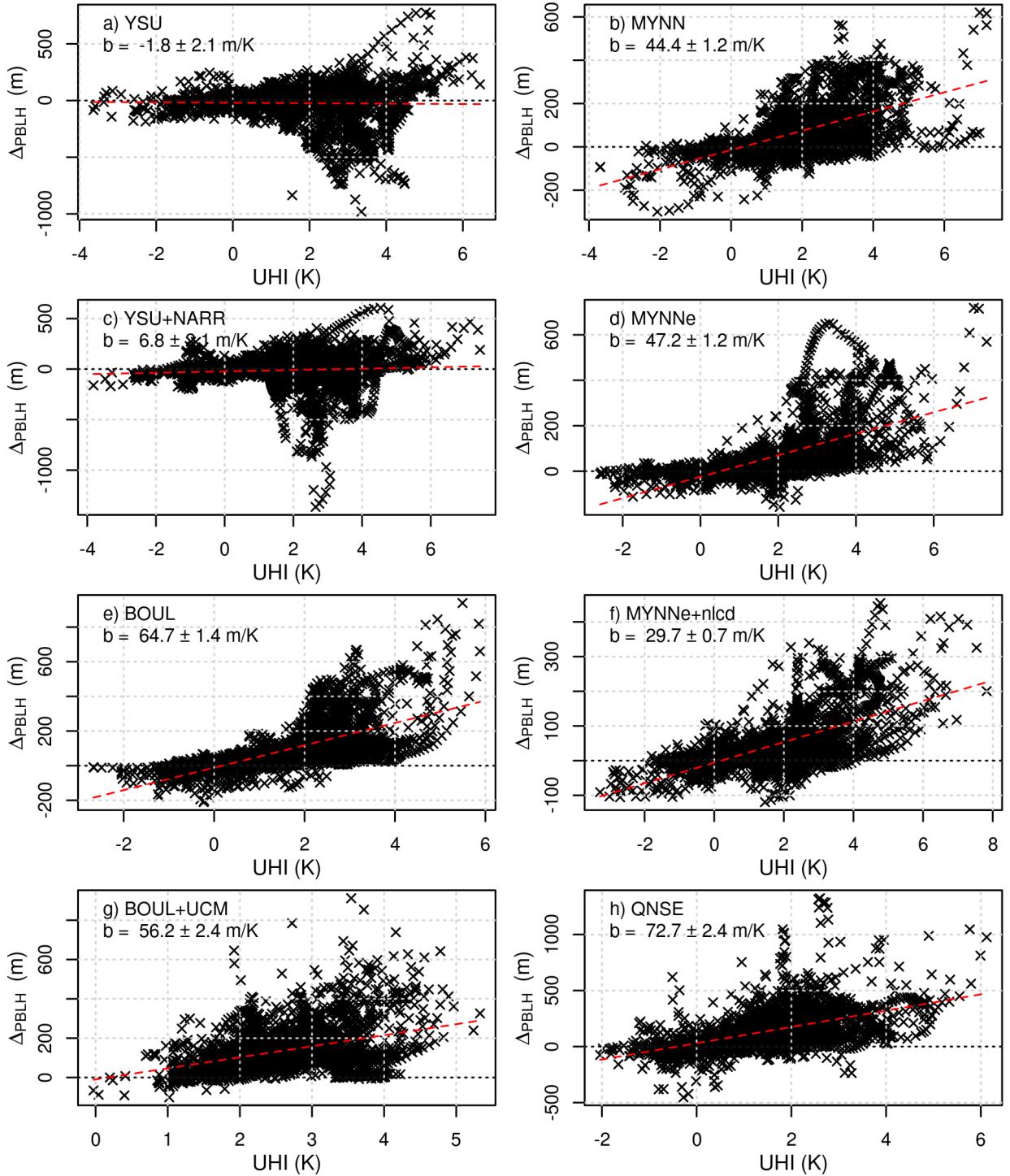
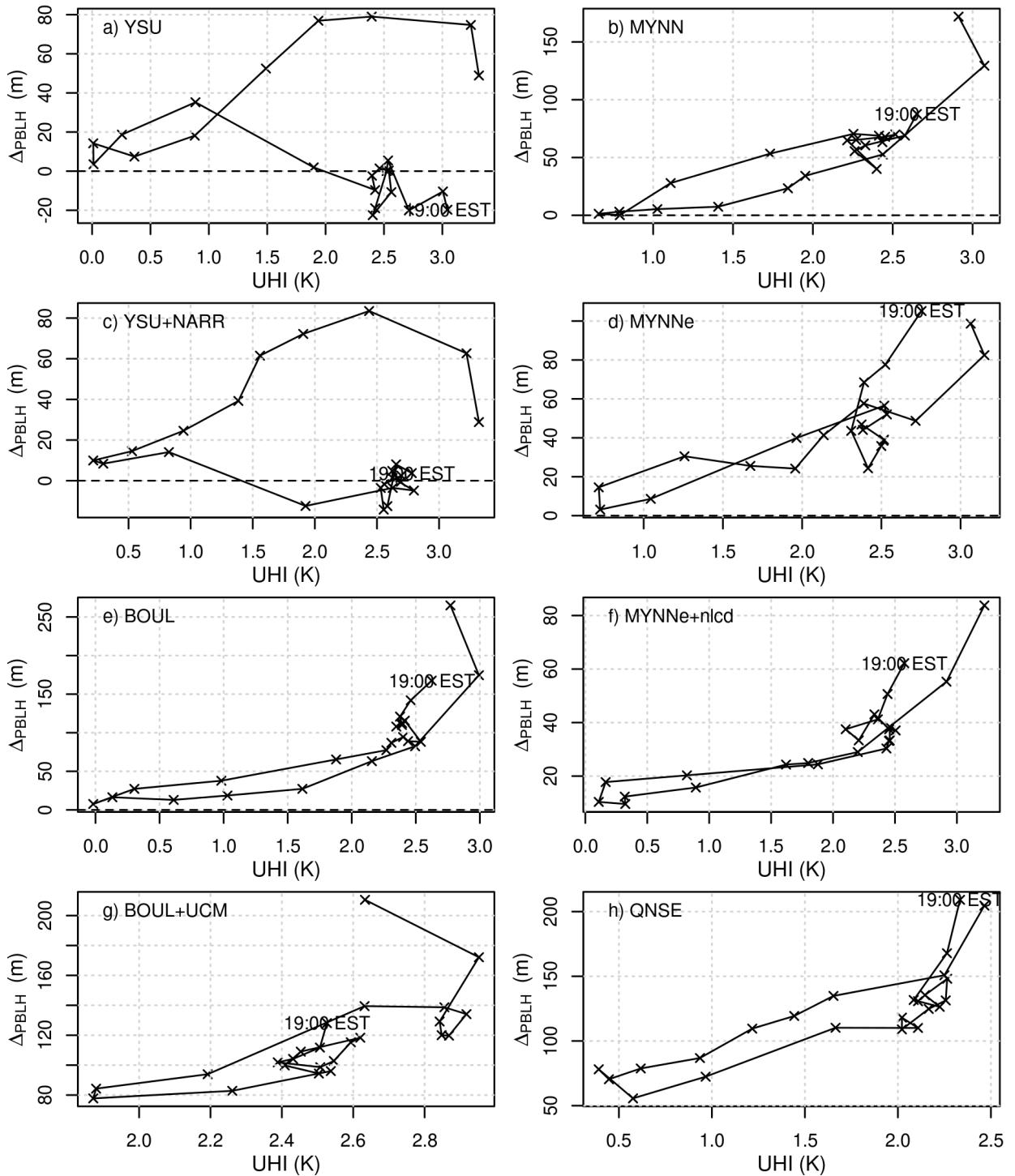
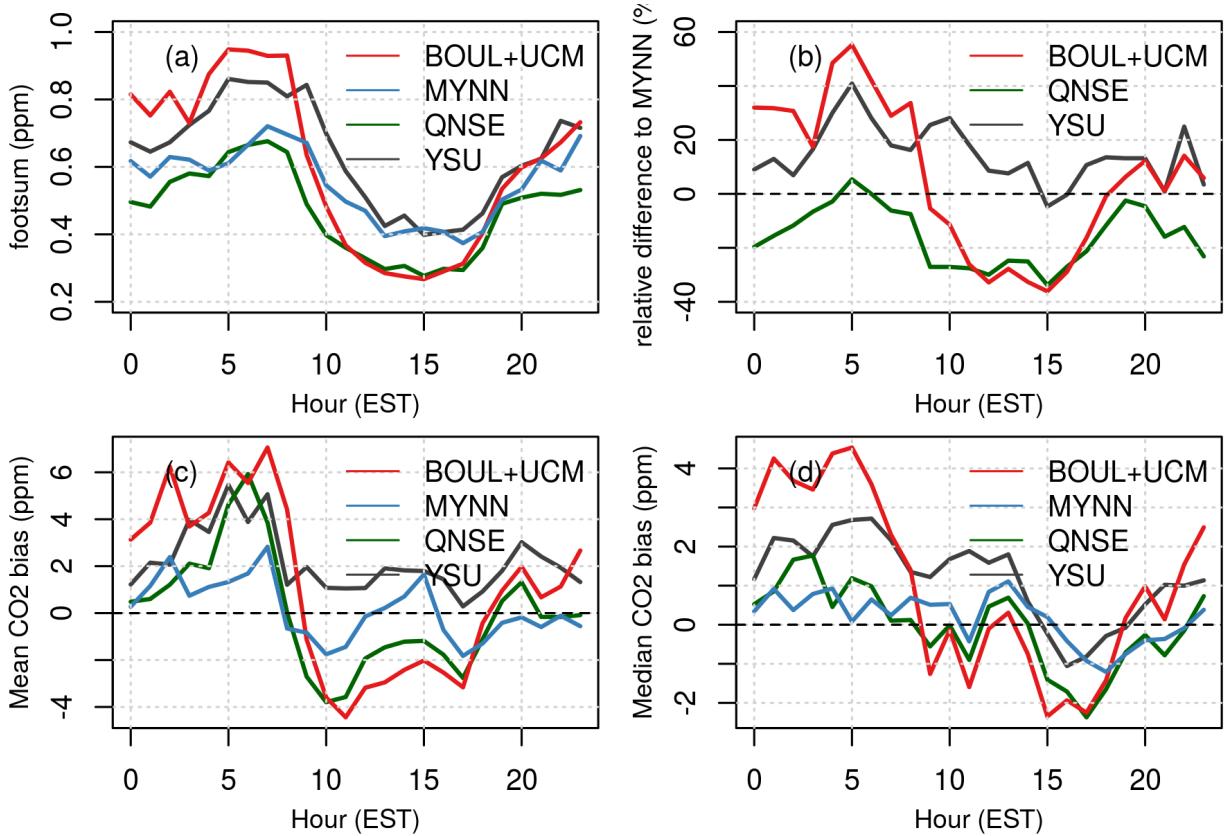


FIG. 8. Scatter plot and linear model fit of the relationship between the UHI and the Δ PBLH. February 2016.



761 FIG. 9. Median daily cycle of the UHI and the PBLH difference between urban and non-urban areas (Δ PBLH).
 762 The marked point represents the time 19 EST while the line connects the points as the time increases, thus ending
 763 at 18 EST. February 2016.



764 FIG. 10. Daily cycle of the a) Integrated footprint (ppm), b) relative difference to MYNN (%),c) CO₂ mean
 765 bias (ppm) and, d) CO₂ median bias (ppm) for the two towers and 4 configurations. February 2016.

Supporting Information for ”Assessment of Planetary Boundary Layer parametrizations and urban heat island comparison: Impacts and implications for tracer transport”

Israel Lopez-Coto¹, Micheal Hicks², Anna Karion¹, Ricardo K. Sakai³, Belay

Demoz⁴, Kuldeep Prasad¹, James Whetstone¹

¹National Institute of Standards and Technology, Gaithersburg, MD

²National Weather Service, National Oceanic and Atmospheric Administration, Sterling, VA

³Howard University, Washington, DC

⁴Department of Physics, University of Maryland, Baltimore County, Baltimore, MD

Contents of this file

1. Figures S1 to S10

Introduction This document contains complementary figures in support of the main text. Figures S1 to S6 show daily cycle of the differences (model - observations) for the Temperature, wind speed, wind direction, planetary boundary layer height, sensible heat flux and latent heat flux respectively while Figures S7 to S10 show the daily cycle of the urban heat island, the area averaged sensible heat flux difference between urban and non-urban areas, the area averaged latent heat flux difference between urban and non-urban

Corresponding author: Israel Lopez-Coto, National Institute of Standards and Technology, Gaithersburg, MD (israel.lopezcoto@nist.gov)

areas and the area averaged Bowen ratio for urban areas over Bowen ratio for non-urban areas.

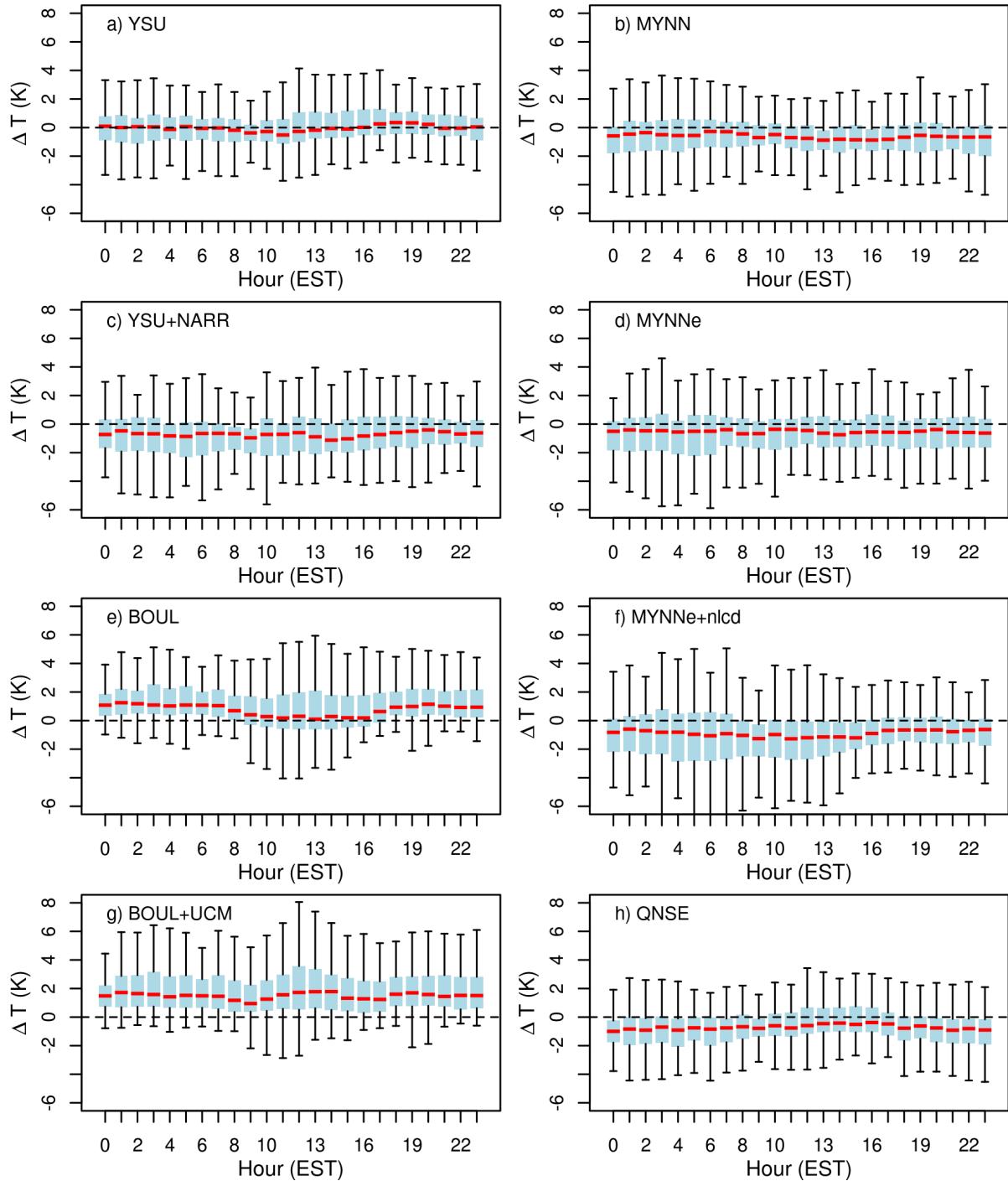


Figure S1. Daily cycle of the differences for the Temperature, including all ISD stations, for the eight configuration tested. Red line is the median, blue bar represents the first and third quantile and the whisker bars are limited to 1.5 times the IQR.

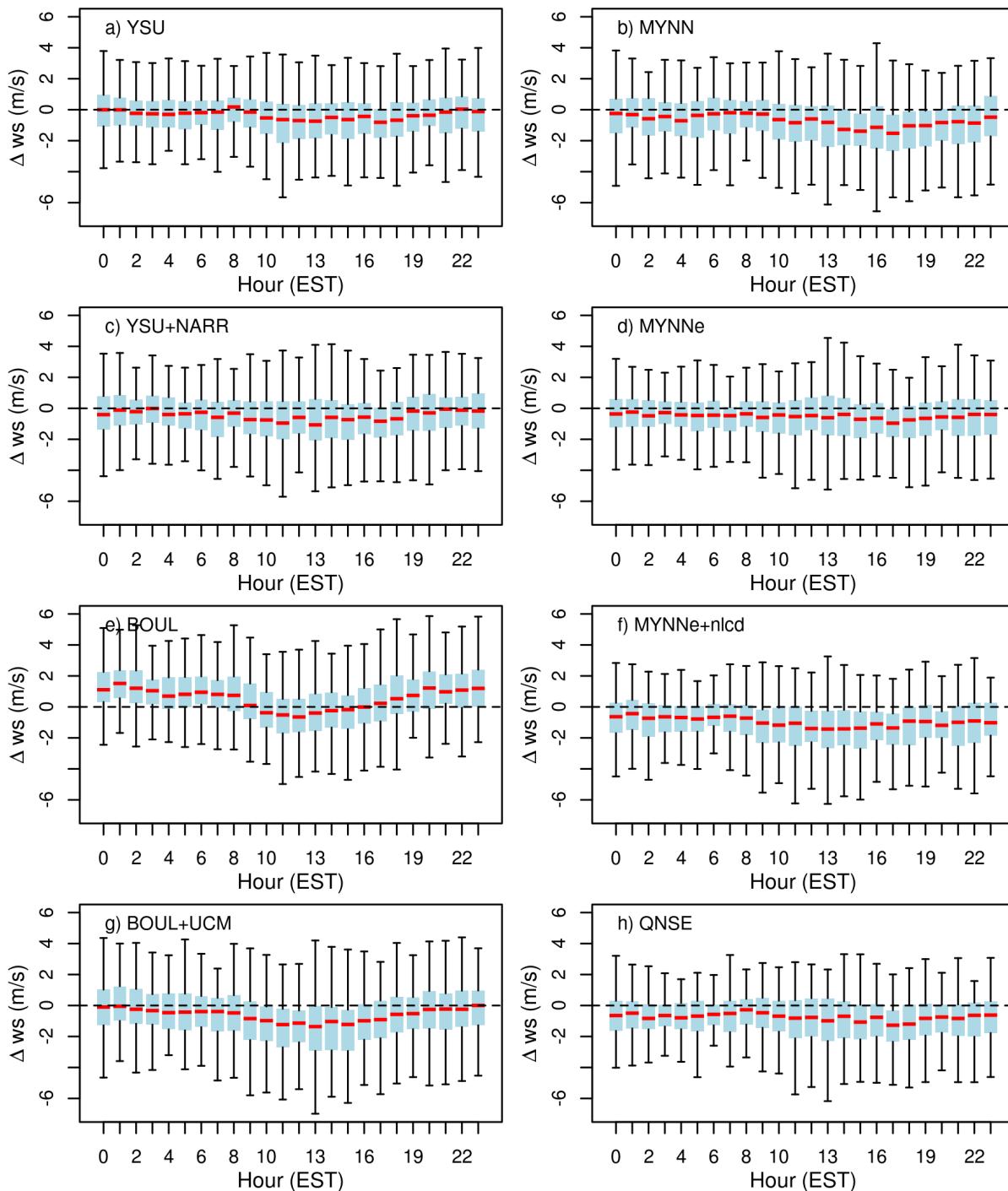


Figure S2. Daily cycle of the differences for the wind speed, including all ISD stations, for the eight configuration tested. Red line is the median, blue bar represents the first and third quantile and the whisker bars are limited to 1.5 times the IQR.

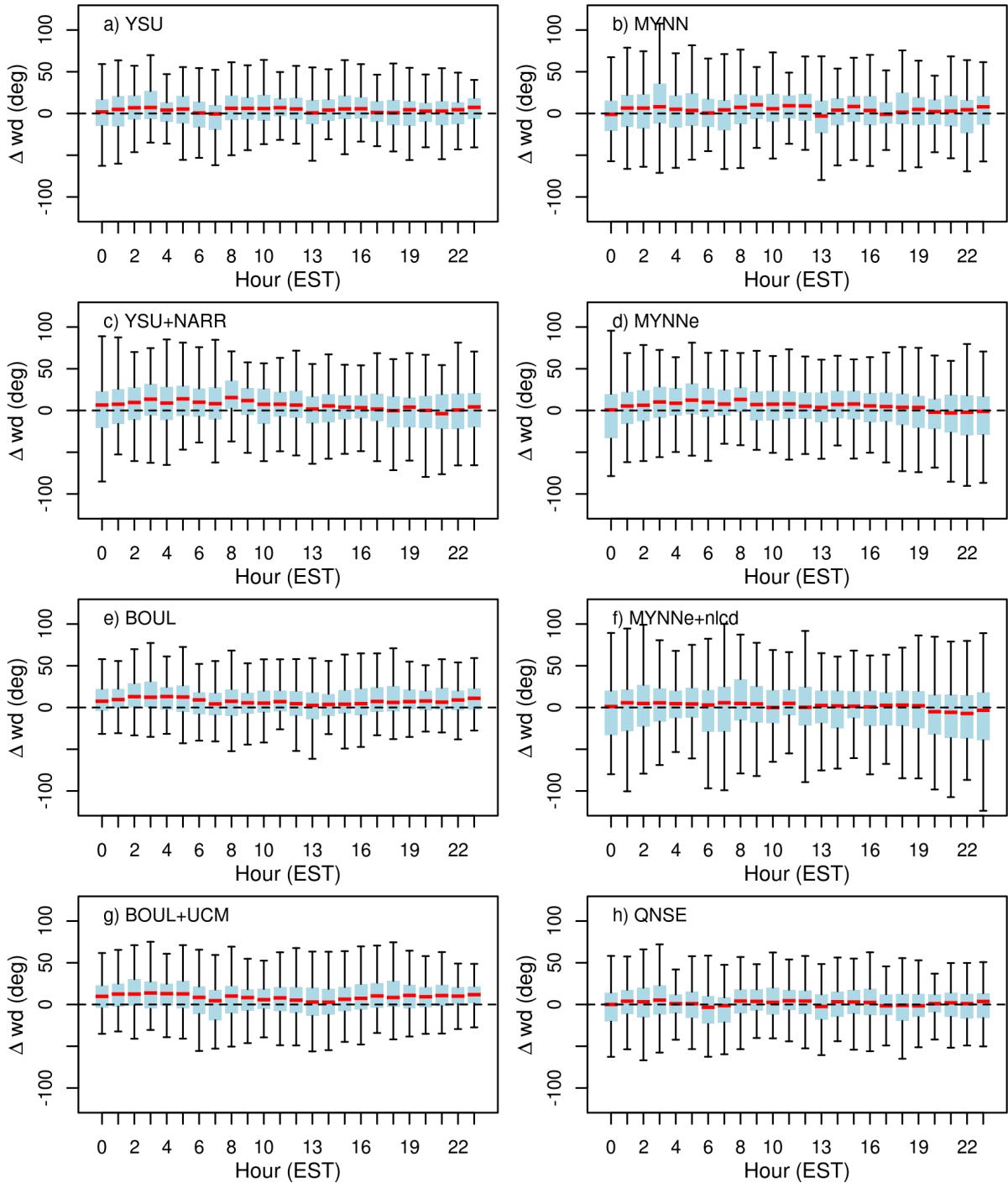


Figure S3. Daily cycle of the differences for the wind direction, including all ISD stations, for the eight configuration tested. Red line is the median, blue bar represents the first and third quantile and the whisker bars are limited to 1.5 times the IQR.

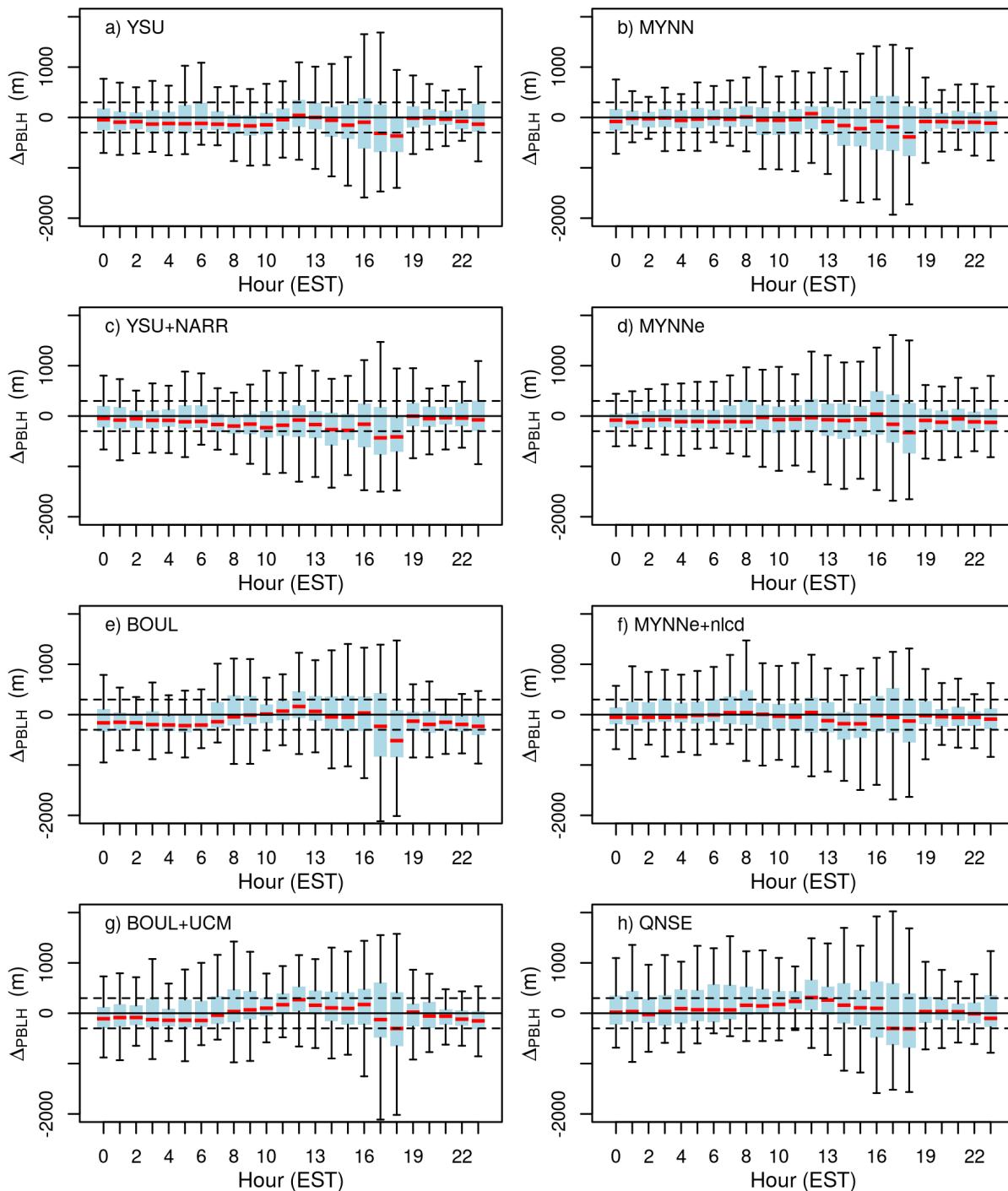


Figure S4. Daily cycle of the differences for the PBLH, including both ceilometers, for the eight configuration tested. Red line is the median, blue bar represents the first and third quantile and the whisker bars are limited to 1.5 times the IQR.

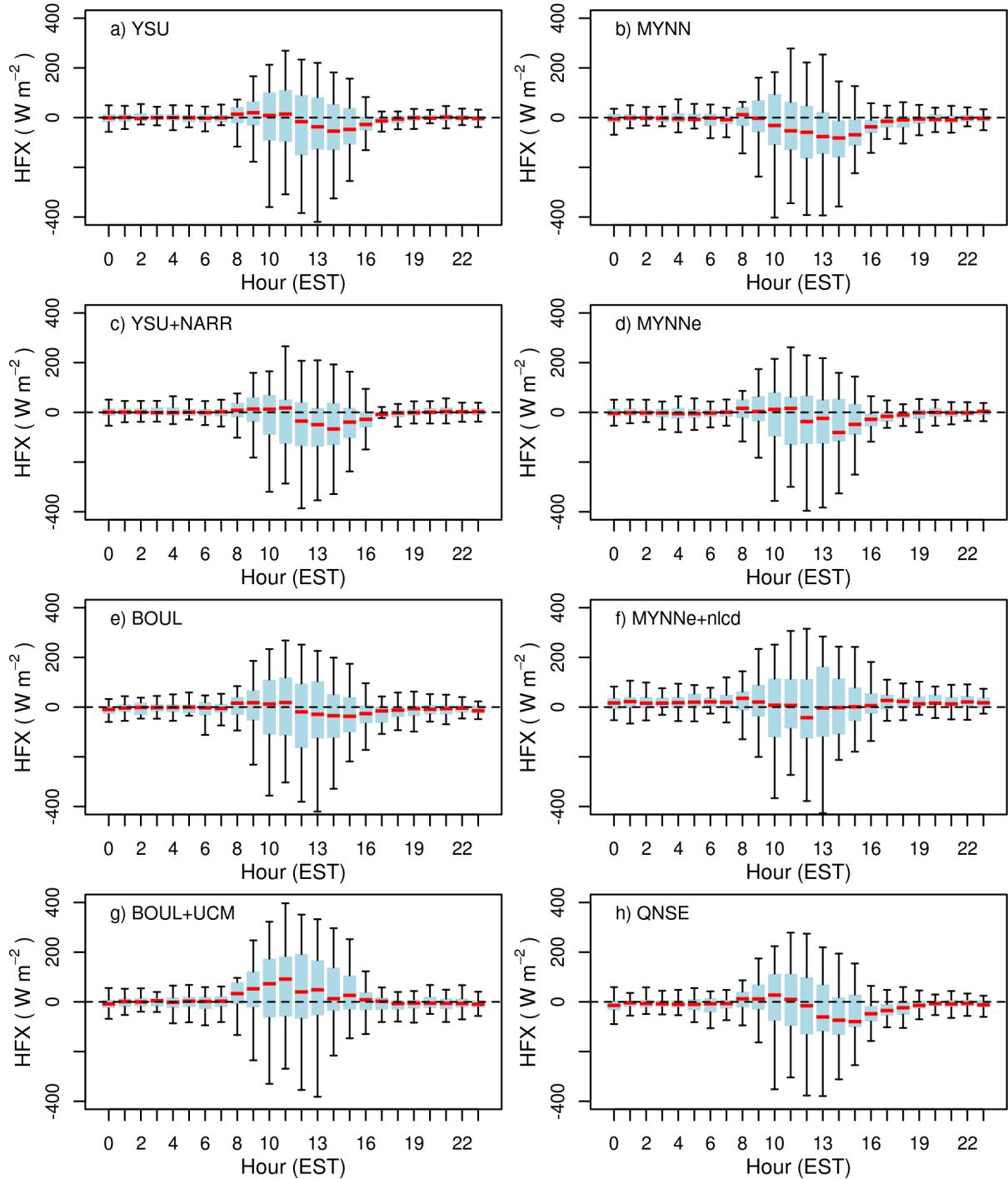


Figure S5. Daily cycle of the differences for the Sensible Heat Flux (HFX) at Betsville for the eight configuration tested. Red line is the median, blue bar represents the first and third quantile and the whisker bars are limited to 1.5 times the IQR.

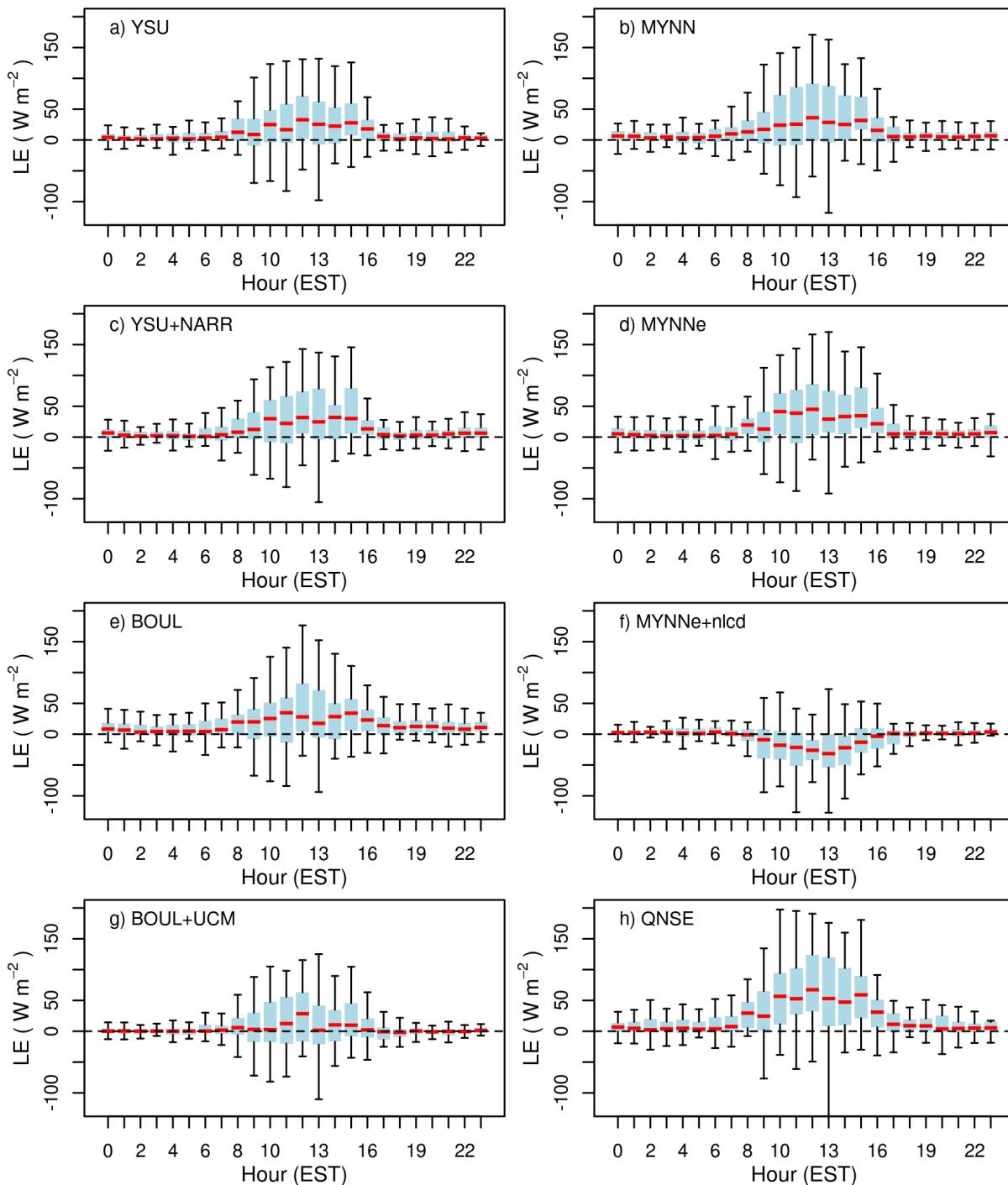


Figure S6. Daily cycle of the differences for the Latent Heat Flux (LE) at Betsville for the eight configuration tested. Red line is the median, blue bar represents the first and third quartile and the whisker bars are limited to 1.5 times the IQR.

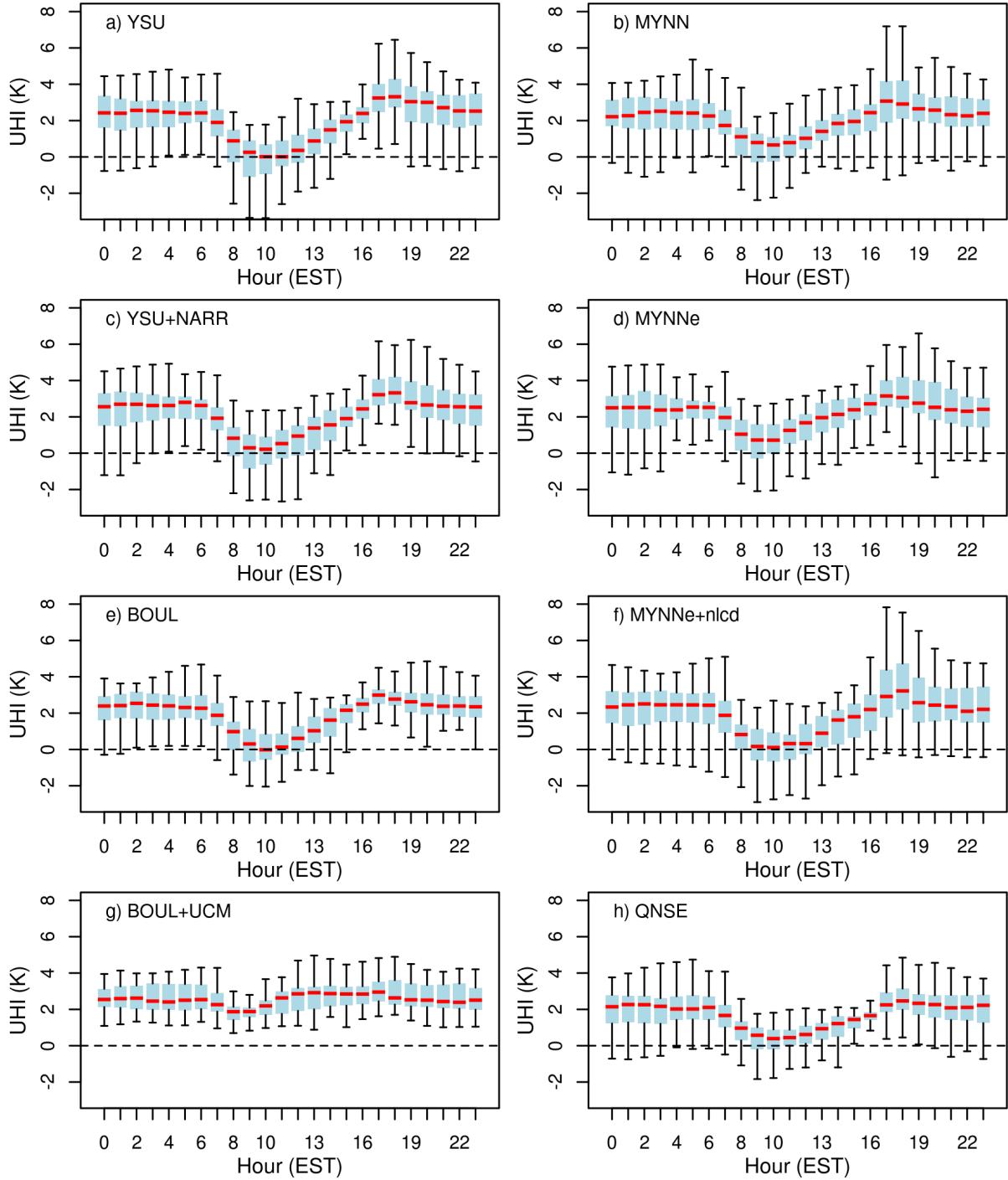


Figure S7. Daily cycle for the Urban Heat Island (UHI) for the eight configuration tested. Red line is the median, blue bar represents the first and third quartile and the whisker bars are limited to 1.5 times the IQR.

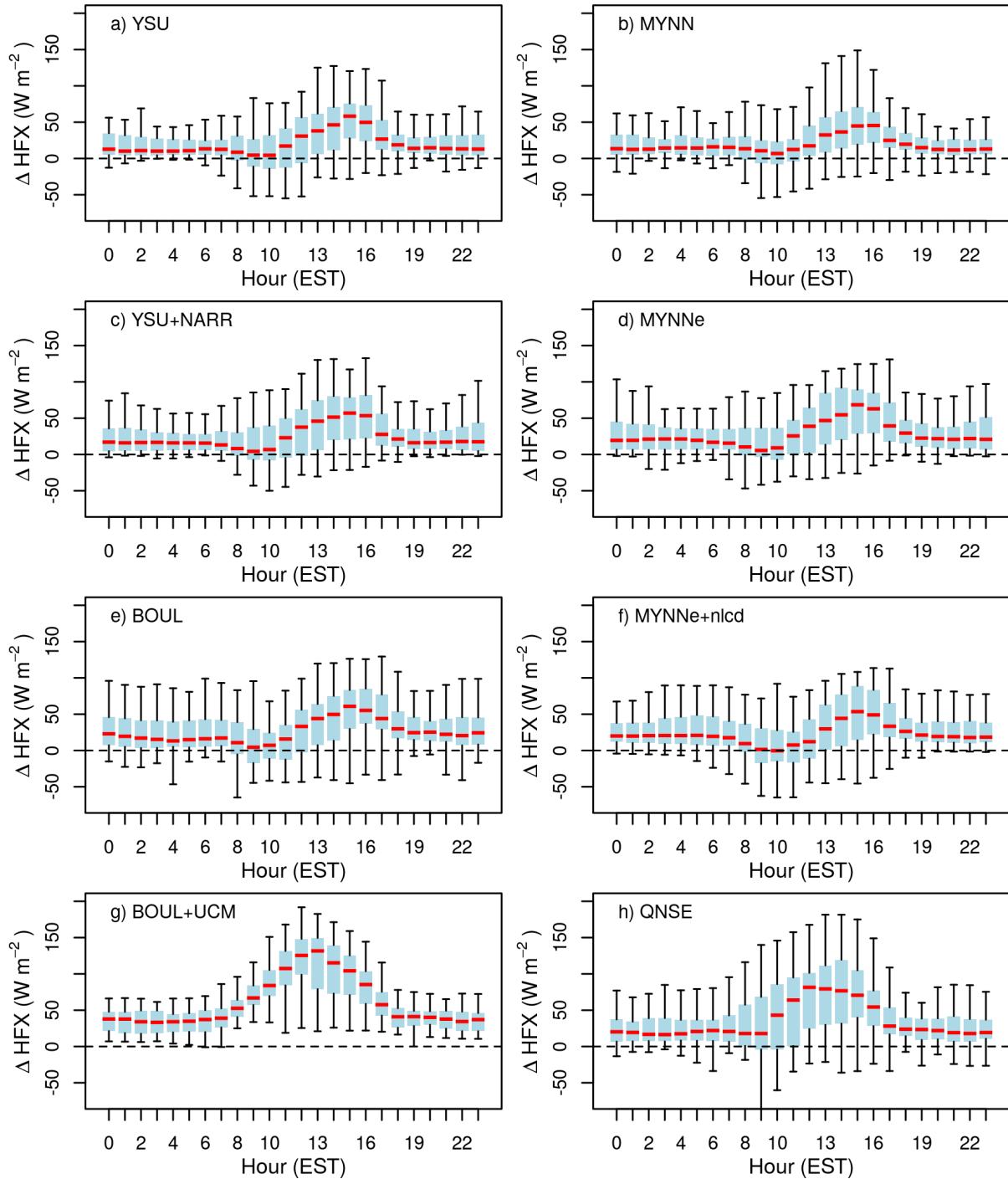


Figure S8. Daily cycle for the area averaged Sensible heat flux difference between urban and non-urban areas (ΔHFX). Red line is the median, blue bar represents the first and third quartile and the whisker bars are limited to 1.5 times the IQR.

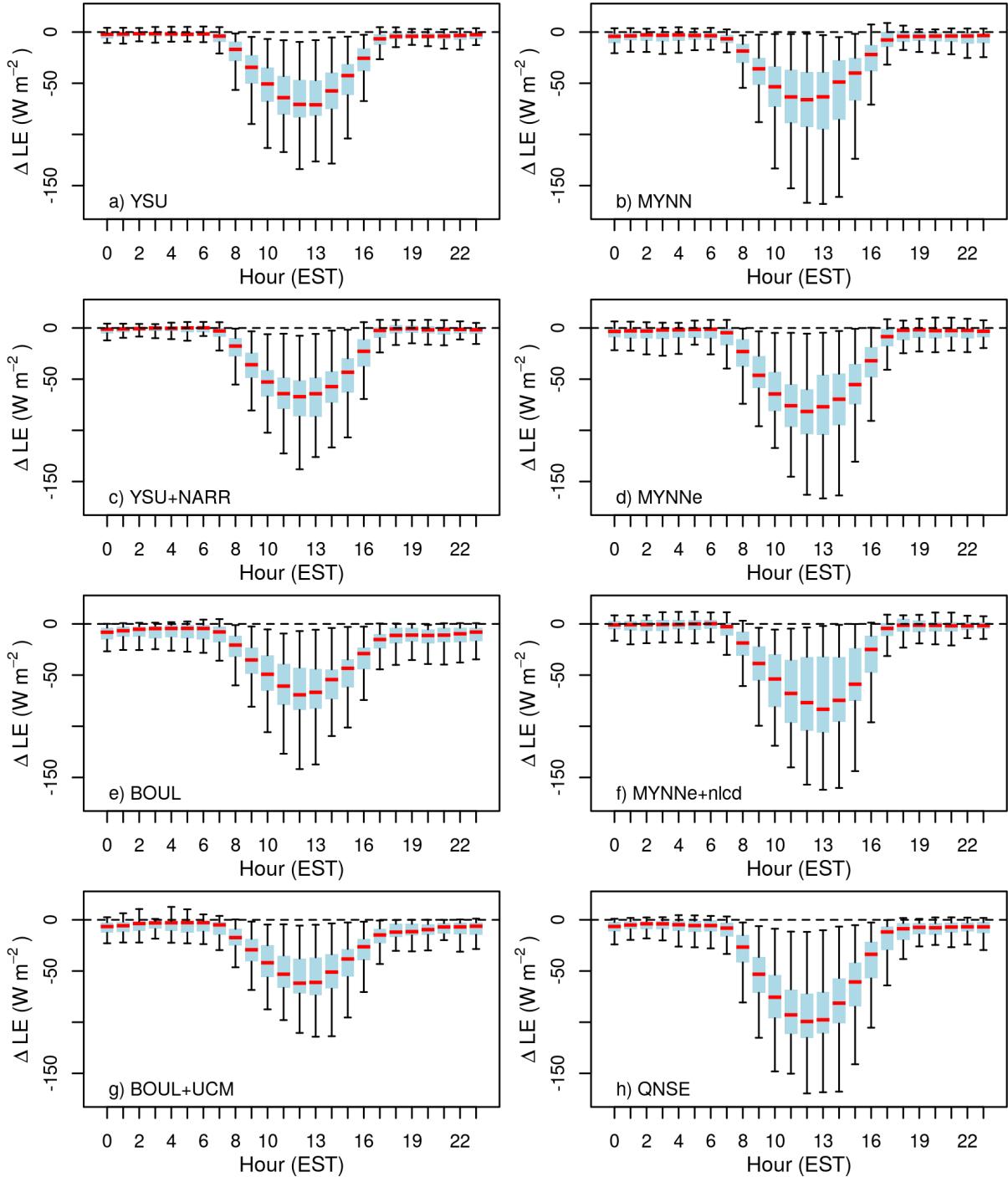


Figure S9. Daily cycle for the area averaged Latent heat flux difference between urban and non-urban areas (ΔLE). Red line is the median, blue bar represents the first and third quartile and the whisker bars are limited to 1.5 times the IQR.

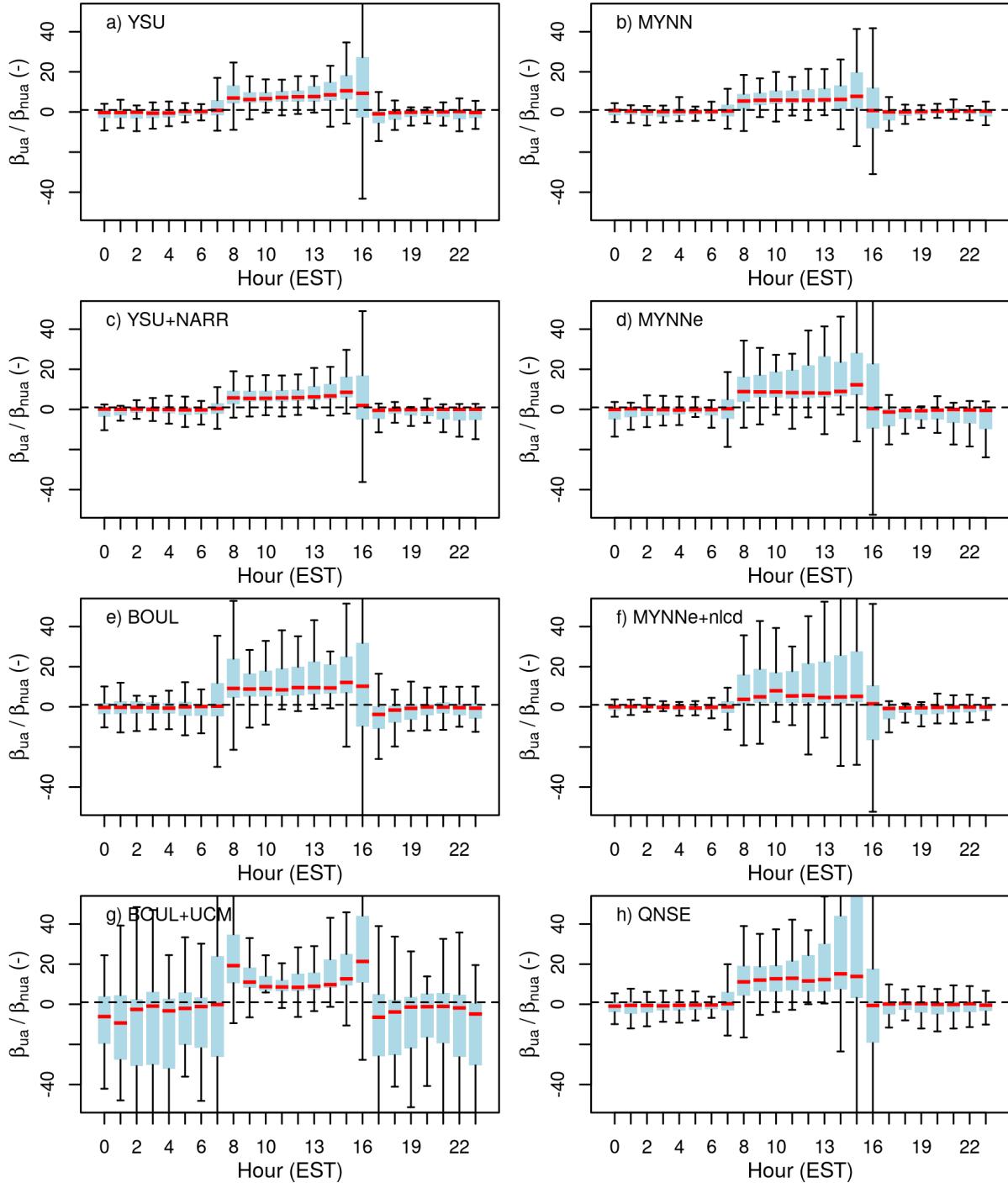


Figure S10. Daily cycle for the area averaged Bowen ratio for urban areas over Bowen ratio for non-urban areas (β_{ua}/β_{nua}). Red line is the median, blue bar represents the first and third quantile and the whisker bars are limited to 1.5 times the IQR. Black horizontal dashed line marks the value 1.